FEASIBILITY STUDY OF

PIEZOTRANSISTOR ACCELEROMETERS

Prepared for

Headquarters
National Aeronautics and Space Administration
Washington, D. C.

Second Annual Report

October 1966

RESEARCH TRIANGLE INSTITUTE, NORTH CAROLINA 27709

FOREWORD

CONTENTS

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

Section I

INTRODUCTION AND DISCUSSION

The purpose of this study has been to determine the feasibility of utilizing the piezojunction effect as the sensory phenomenon in accelerometers and other transducers. The present report covers the second year of study. The major features of the piezojunction phenomenon were reported on in the first annual report in September 1965 (hereafter referred to as I). The results reported there will not be repeated in detail here. The reader is, therefore, referred to the first annual report in order to appreciate the total results of this study. Although the present report considers some of the more fundamental concepts of the phenomenon, it is primarily concerned with the applications of the phenomenon and in particular the development of the silicon needle sensor.

As shown in I, large mechanical stress applied to the junction area in silicon and germanium p-n junction devices can cause large reversible changes in the electrical properties of the devices. In diodes, the junction current, at a constant applied voltage, can be made to increase by orders of magnitude. In transistors, stress causes a large decrease in gain. Four-layer switches experience a decrease in break-over voltage with stress.

The effect of stress on four-layer switches has been considered in some detail since it is particularly attractive for use in transducers which give a digital output as shown in I.

A theoretical discussion of the effects of mechanical stress on the electrical characteristics of p-n-p-n diodes is given in

Appendix A. It is shown there that if the total junction area of the switch is stressed, the break-over voltage decreases. If only a small region of the switch is stressed, the break-over voltage can increase. Both of these effects have been verified experimentally, however, due to the complexity of the problem, it has not been possible to make any meaningful direct comparison with the theory. A pressure transducer based on a p-n-p-n switch in a relaxation oscillator is discussed in section 4.4.

Although direct stressing of a four-layer switch is desirable in many cases, the same effect can be obtained by using a stressed diode to bias the gate in commercial four-layer switches. Relaxation os illators based on this principle are discussed in Section V.

The effects of mechanical stress on the breakdown voltage of p-n junctions is discussed in Appendix B. (Appendix B is a copy of a paper published in the September, 1966, issue of The Journal of Applied Physics.) It is shown there that the breakdown voltage of silicon diodes is reduced by mechanical stress. The sensitivity of the breakdown voltage to stress is much smaller than for example the current in diodes. The relative change in breakdown voltage is given by

$$\frac{\Delta V}{V_B} \simeq - (10^{-12} \text{ cm}^2/\text{dyne}) \ \sigma \ , \qquad (1.1)$$

where σ is the magnitude of the applied stress. The effect is shown to be independent of crystal orientation. In germanium, the effect is a much more complicated function of the stress. For low stress

levels $\Delta V/V_B$ increases while a large stress level gives a decrease in $\Delta V/V_B$.

To date, the effect of stress on the breakdown voltage has not been exploited in any transducer applications. In order to utilize the effect, one need only bias the diode to be stressed in the breakdown region. If the voltage is held constant across a device which has a hard breakdown, a small change in the breakdown voltage will result in large changes in the junction current. Since the breakdown voltage of p-n junctions is relatively insensitive to temperature changes, the use of this mode of operation in p-n junction sensors could prove very fruitful.

Due to the large stress levels required to cause the piezojunction effect, there have been only two methods used to stress the junctions in transducer applications. The first of these is to couple the mechanical stress to the desired junction area by means of a diamond, sapphire, or steel needle or indenter. This method was the first to be used and has the attribute of allowing complicated junction devices to be stressed. Some applications which utilize this method are described in the following sections of this report. The second method for applying the large stress levels was developed during the present study and was reported in I. This method is the inverse of the indenter method in which the p-n junction or junctions to be stressed are formed on the apex of a silicon needle (silicon needle sensor). Much effort has been placed on this approach and its use in transducers during the past year of study. These results indicate that the approach is sound and it shows great promise for use in a variety of transducer applications.

A variety of transducers and configurations have been investigated, however, the major emphasis has been on the demonstration of the use of the piezojunction sensor in accelerometers. A dual diaphragm accelerometer which uses a silicon needle sensor has been developed and tested. This accelerometer can be made to operate over a large range of accelerations with both ac and dc. It is insensitive to cross axis acceleration and can have rather large resonant frequencies.

Methods for achieving digital outputs in piezojunction sensors have been considered as well as analogue outputs. Since the piezojunction sensors are, themselves, solid state devices, they are, of course, ideal elements to use with conventional microelectronic techniques.

Included in Appendix C is a copy of a paper which was published in the August, 1966, issue of the Journal of Applied Physics which describes the effects of mechanical stress on the generation-recombination current in p-n junctions.

Section II

SILICON NEEDLE STRESS SENSOR

2.1 Introduction

It has become evident during the course of this study that the "semiconductor needle" approach to utilizing the piezojunction phenomenon in transducers has many advantages over other known methods. In this approach, the sensory junction or junctions are formed on the apex of a semiconductor needle as shown in Fig. 2-1. The apex of the needle is pressed against a smooth hard surface to create the desired mechanical stress in the junction or junctions. The magnitude of the mechanical force on the needle causes the electrical characteristic of the junction or junctions to change resulting in the desired transducer action.

The electrical and stress characteristics of needle transducers not only depend on the quality of the p-n junctions but also depend on the geometry and surface roughness of the needle point. In order to fabricate, in a reproducible manner, needle sensors, one must be able to control the needle geometry. Work on this phase of the present study has led to the following fabrication techniques.

The needle sensor is formed by: (1) fabricating the desired silicon blank (rectangular rod), (2) cleaning the blank, (3) electro-etching a conically shaped point on the blank, (4) cleaning, and (5) forming the p-n junction or junctions on the apex of the needle.

2.2 Needle Fabrication Method

The first step in the needle fabrication process is to form the needle blank. The blank is a rectangular bar of silicon approximately

(5)
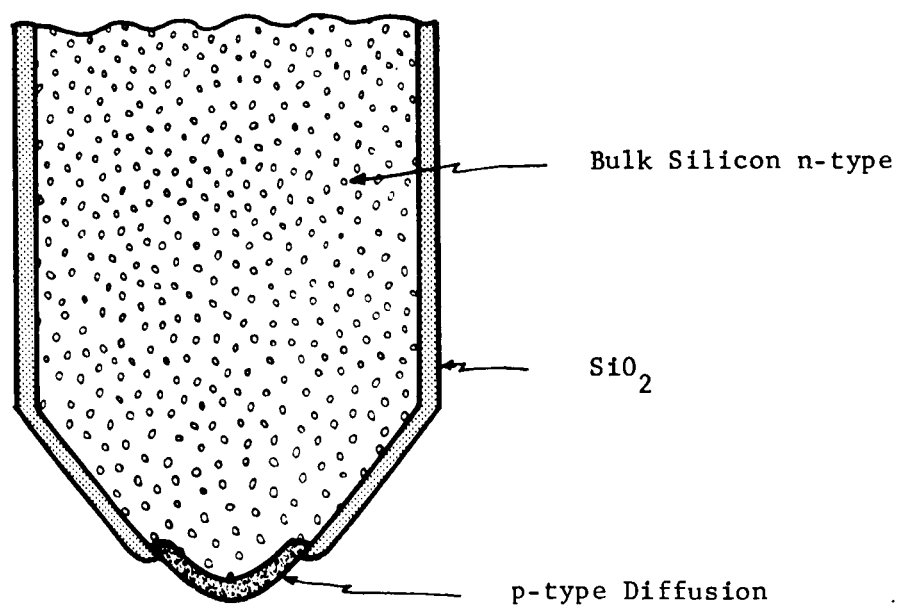
Fig. 2-1. A Sketch of the Silicon Needle Diode Sensor.

1/2 inch long and 20 mils on each side. The dimensions of the bar are not important with the exception that the sides must be very nearly equal (square perpendicular to the axis). If the rod is not square to within 1/2 mil on each side, the resulting needle tip will be wedge shaped and not conical as desired. A round rod is more desirable than a square rod, however, experiments have shown that the square bars are acceptable.

A cleaning process similar to that used commercially in the production of semiconductor devices was modified and used to clean the silicon starting blanks in order to insure proper etching. This process in step form is as follows:

(1) Ultrasonically clean in TCE for 2 minutes.

(2) Rinse in TCE.

(3) Clean in Acetone for 5 - 10 minutes.

(4) Rinse in de-ionized water.

(5) Dry on filter paper.

(6) Plate in $CuSO_4$, HF solution for 5 minutes.

(7) Rinse two times in de-ionized water.

(8) Rinse in HNO_3 (70%) 3 times.

(9) Rinse in de-ionized water.

(10) Boil in 30% H_2O_2 for 10 minutes.

(11) Rinse in de-ionized H_2O.

(12) Ultrasonically clean in de-ionized H_2O 5 minutes.

(13) Rinse 3 times in de-ionized H_2O.

(14) Rinse in Acetone 2 - 4 minutes.

(15) Dry on filter paper.

(16) Rinse in HNO_3 at 50° C.

The purpose of this procedure is to thoroughly remove all contaminants from the surface of the silicon, including any that might have been introduced during the mechanical slicing operation.

The next step in the needle formation process is the electrochemical etching of the needle tip. This was done by slowly lowering and raising one end of the needle blank into the etch solution. An electric potential (a.c.) was maintained between the needle blank and an electrode in the etch solution. A motor driven apparatus was designed and assembled which controlled the amount of time the starting bar remained in the etching solution and the depth to which it is submerged. It is desirable that this depth be as shallow as possible in crder that the mechanical strength of the final needle be maximized. The shallowest depth practical was found to be approximately 1/8 of an inch. The silicon bar was submerged into and retrieved from the etching solution by a piston driven by the motor through a lever arm. A photograph of the apparatus is shown in Fig. 2-2.

The solution used to etch the rectangular silicon bars into needles was composed of a mixture of concentrated HNO_3 and concentrated HF. The mixture ratio was defined as volume of HNO_3 divided by volume of HF and was varied from a high of 14 to a low of 3.5 in an effort to determine its effect on the completed needle. Simultaneously, the electroetching current was varied from 20 ma to 70 ma. The purpose of this procedure was to obtain a balance between chemical and electrical etching phenomena. It was determined that a mixture ratio above 7 was not sufficient to provide a proper chemical contribution to the etching process while a ratio below 4.5 provided an excessive chemical contribution. A low ratio caused an "orange
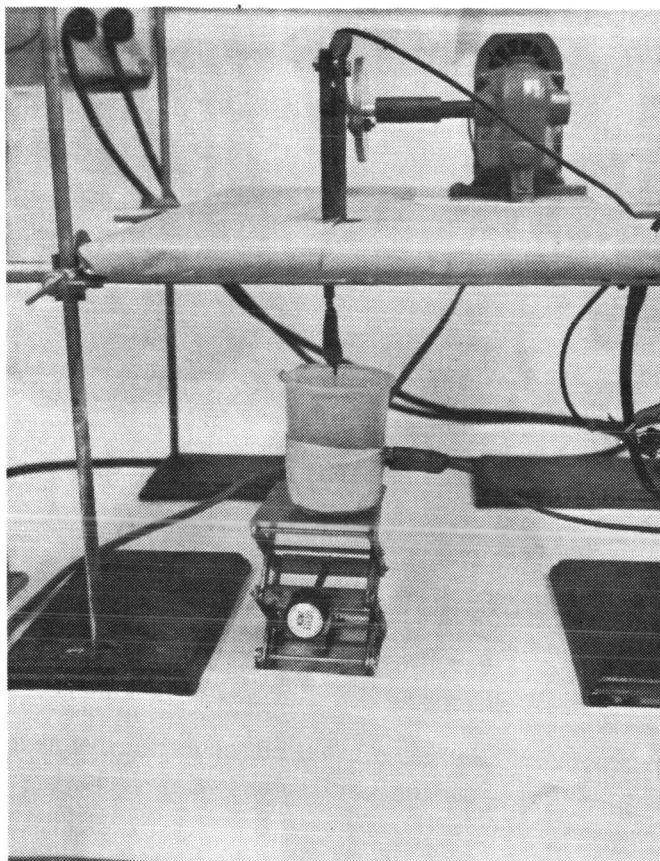
Fig. 2-2. Photograph of the Apparatus Used to Etch the Silicon Bars
Into Needles.

peel" effect on the surface of the needle tip and caused the cone of the

needle to have an uneven profile. An example of an excessively low

mixture ratio is shown in Fig. 2-3. A high mixture ratio resulted in

blunt tips because etching action was not sufficient to effectively

sharpen the silicon. An example of a high ratio is shown in Fig. 2-4.

When the mixture ratio was outside the prescribed bounds, a proper

balance could not be attained by manipulating the current level. A

current level above approximately 60 ma increased the electrical contri-

bution to a point of unattainable equillibrium while current below 20 ma

resulted in an unbalance in favor of chemical etching. The optimum

mixture ratio-current combination was found to be approximately 4.7 at

10
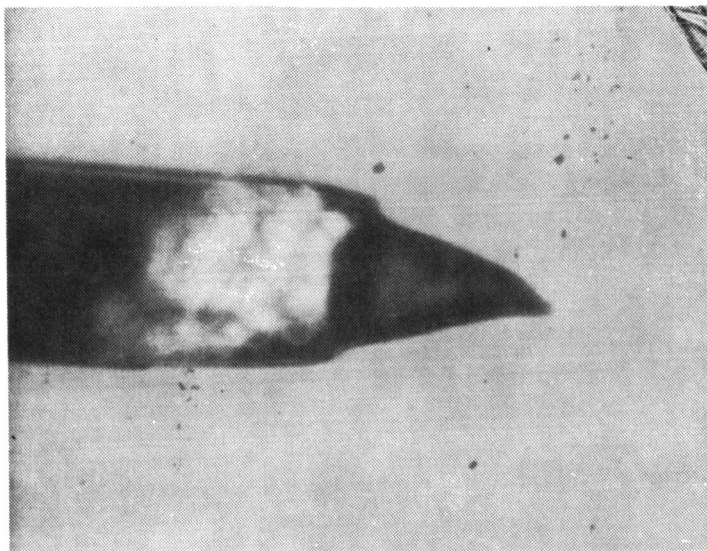
Fig. 2-3.   Photomicrograph of a Silicon Needle - Low Mixture Ratio
            (1 Ω-cm n-type silicon).
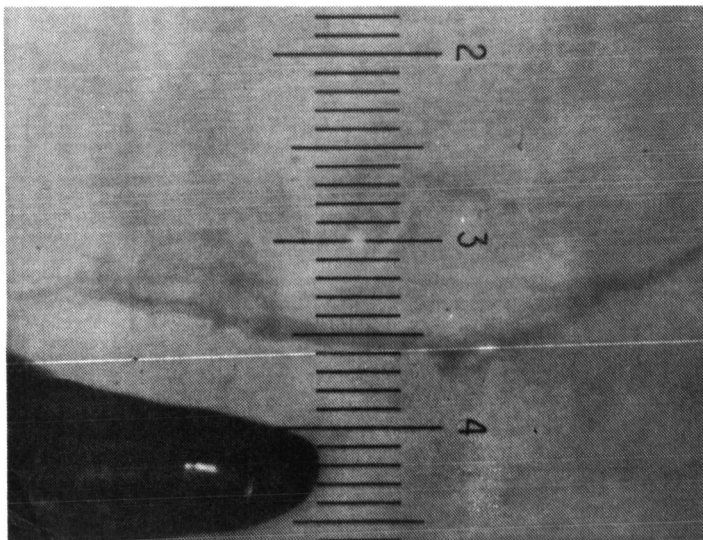
Fig. 2-4.   Photomicrograph of a Silicon Needle - High Mixture Ratio
            (1 Ω-cm n-type silicon).

50 ma. However, the current level did not have nearly as much affect on the needle as did the mixture ratio.

Also worthy of note is the manner in which the solution was mixed and which had a significant effect on the resultant needle. A non-uniformly mixed solution produced needles of various uncontrollable geometries with all other parameters fixed. This problem was allevi-ated by agitating the mixture for a period of 10 to 15 minutes before the electroetching process was initiated.

The parameters of time and dip-rate were found to be equally as important as the mixture ratio and current parameters. The dip-rate parameter is the rate at which the silicon blank is lowered into and withdrawn from the etching solution. The time parameter is the elapsed time from the beginning to the end of the dipping process. An ex-cessively short time provided insufficient etching and left the needle with a concave cone and a blunt tip. An excessively long time provided excessive etching and left the needle with a convex cone and a blunt tip. The optimum time was found to be approximately 4 minutes. An example of excessive time is depicted by the needle tip in the photo-micrograph of Fig. 2-5 while the effect of insufficient time can be seen in Fig. 2-6.

The dip-rate was found to have a detrimental affect on the finished needle only if this rate was very low. At low dip rates, an "orange peel" effect was present on the surface of the etched portion of the needle. This was attributed to chemical etching during the up portion of the dip rate cycle. Chemical etching was caused by the fact that some residual solution was withdrawn along with the needle, and at low dip-rates, had sufficient time to react while the needle was out of
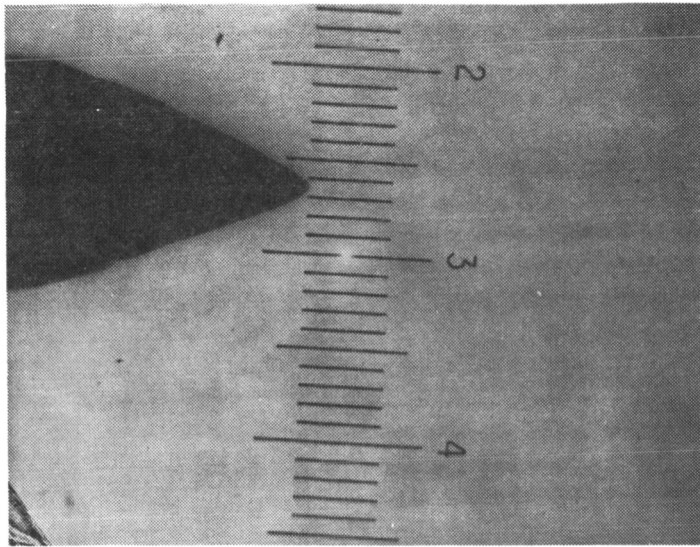
Fig. 2-5. Photomicrograph of 1 Ω-cm n-type Silicon Needle - Etching
Time 6 min. (each division is .04 mm).
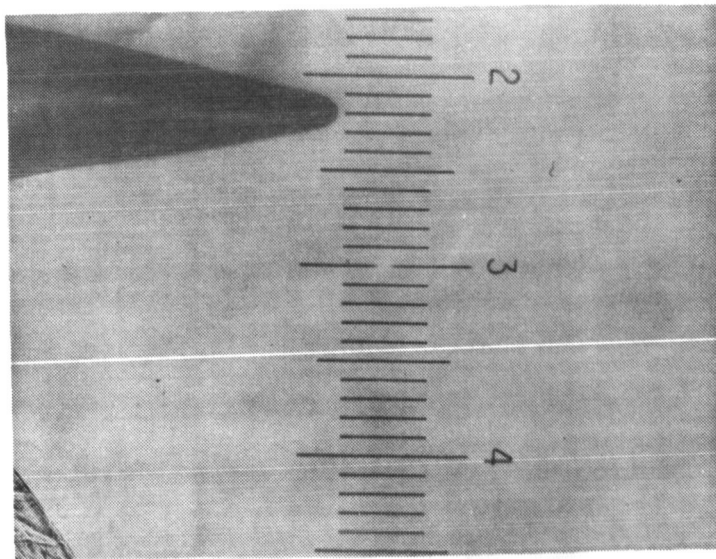
Fig. 2-6. Photomicrograph of 1 Ω-cm n-type Silicon Needle - Etching
Time 3 min. (each division is .04 mm).

the solution. Since the needle was out of the solution and thus no current could flow, the etching action was all chemical during this portion of the cycle. The "orange peel" effect was eliminated by increasing the dip-rate to approximately 20 cycles per minute.

The influence of the various parameters such as mixture ratio, etching current, time and dip-rate on the final needle are summarized in Table I. Table I was obtained for 1 ohm-cm n-type silicon. However, the only parameter appreciably affected by a change in resistivity is the current level. A higher current is required for lower resistivities material and vice versa. No appreciable effect has been noted for a change from n to p-type silicon. For very high resistivity starting blanks, an ultraviolet light source aids the etching process.

Using the results of these experiments, it has been possible to reproducibly fabricate silicon needles with approximately the desired geometrical properties. The radii of curvature, as well as tip lengths and surface characteristics, can be controlled to the necessary tolerances. Two needles made using the optimum parameters of Table I are shown in the photomicrographs of Fig. 2-7. It should be noted that the motor driven apparatus is capable of etching any number of needles simultaneously.

2.3 Junction Formation

Once the needle fabrication process was perfected, the next logical step in the development of improved needle sensors was the junction fabrication process. An effort was made to control the parameters of this process with primary emphasis on the method of junction area definition.

14

Table I

Influence of Electroetching Parameters on Silicon Needle Fabrication

Mixture Ratio	Current (ma)	Time (min)	Dip-Rate (cpm)	Material	Resistivity (Ω - cm)	Remarks
14	70	3	7.7	n-Si	1	Insufficient etching
14	70	3.5	7.7	n-Si	1	Smooth surface, rough concave profile
14	70	4.0	7.7	n-Si	1	Same as above
9.3	70	3.0	7.7	n-Si	1	Rough profile, smooth surface, blunt tip
9.3	65	3.5	7.7	n-Si	1	Concave profile, smooth surface, blunt tip
9.3	65	4.0	7.7	n-Si	1	Slightly concave profile, smooth surface, blunt tip
7	65	4.0	7.7	n-Si	1	Smooth surface, rough profile, tip relatively sharp
4.7	65	5.0	7.7	n-Si	1	Profile slightly convex, tip sharp, "orange peel" surface
3.5	65	5.0	7.7	n-Si	1	Blunt tip, profile convex, "orange peel" surface
4.7	50	4.0	7.7	n-Si	1	Sharp tip, profile rather rough, "orange peel" surface
4.7	50	4.0	20	n-Si	1	Sharp tip, smooth profile, surface very smooth
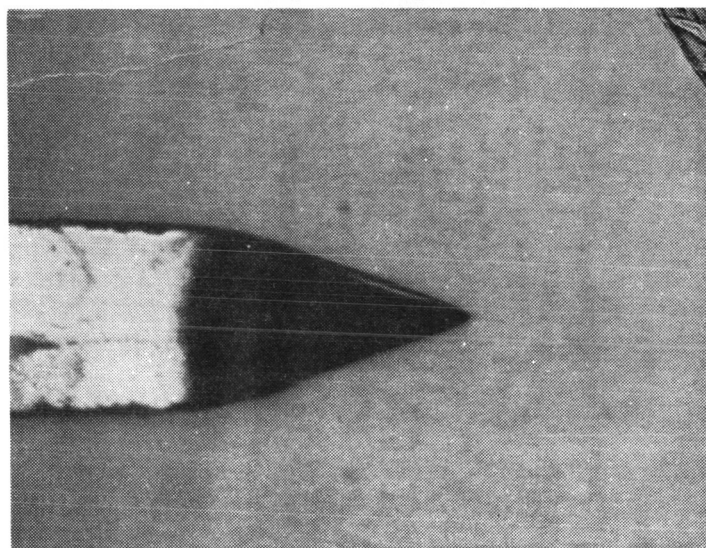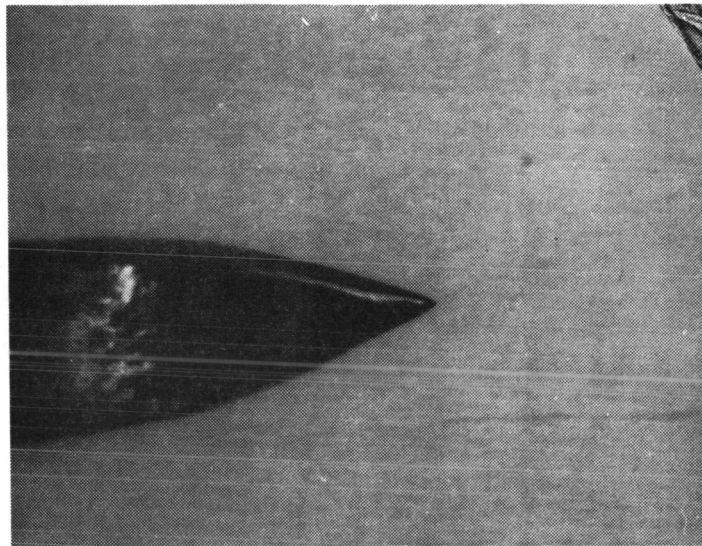
Fig. 2-7. Photomicrograph of Two Silicon Needles Fabricated Using Optimum Parameters of Table I (1 Ω-cm n-type silicon).

The stress sensitivity of p-n junctions depends, among other factors, on the ratio of stressed to unstressed junction area. Therefore, it is desirable that as much of the junction area as possible be exposed to the applied stress. The insensitivity to stress noted in some of the early needle sensors was attributed to the fact that the junction area on the tip of the needles was large with respect to that portion of this area which was actually stressed. An effort was undertaken to develop a technique of junction definition which would reduce the total junction area to that actually stressed.

The present study has been confined to the planar process for fabricating junctions on the tip of the needles. Planar needle sensors were originally fabricated by first growing an oxide on the entire silicon needle. The oxide covered needle was then coated with photoresist (such as KTFR) which remains intact after the exposure and developing process. To expose the tip of the needle, it was carefully lowered into molten black wax which defined the junction area and was then exposed to ultraviolet light. Since the area of the junction was determined by the depth to which the needle tip was lowered into the protective wax, this was the most critical and most difficult step to control in the process. It was most difficult to confine the wax to the very tip of the needle and even more difficult to mask two needles identically. Although this masking technique was somewhat successful, the need for a more precise method led to the development of other techniques.

The first technique investigated in an attempt to obtain a junction only on the portion of the needle to be stressed consisted of lowering the photoresist coated needle into a glass beaker until the needle tip

came into contact with the bottom of the beaker. The beaker was then filled with black ink for the purpose of attenuating the exposure light. The idea here was that only the very tip of the needle would touch the glass and that only this portion of the photoresist would be exposed when light was passed through the bottom of the beaker. This method did not prove to be successful and was, therefore, abandoned.

The next attempt consisted of lowering the photoresist coated needle onto a spring loaded metal plate in much the same manner as if stress was being applied to a finished sensor. The metal plate was coated with black paint so as to protect the portion of the needle in contact with the metal from the exposure light. Once contact was made, the unprotected photoresist on the needle was exposed with an ultra-violet light source for a period of approximately 2 minutes.

The needle sensors produced by the latter method were very sensitive to stress and diodes formed by using this process possessed very good electrical characteristics. Good needle sensors have been obtained by this procedure using both n and p-type silicon starting material of various resistivities. Some of the advantages of this technique over the black wax method are (1) smaller junction area definition and (2) more controllable junction definition. This technique is also much easier and faster than any other method tried to date.

Following the exposure operation, the photoresist was developed leaving a hole in the photoresist on the apex of the needle. The needles were then placed in HF to etch a hole in the oxide on the un-protected needle tip. Care must be exercised in the etching operation to insure that the photoresist does not peel off. This is best accomplished by using a pilot wafer on which the etch time is determined.

A p-n junction can be formed on the apex of the needle by simply removing the photoresist and placing the needle into an appropriate diffusion furnace. Single junctions with excellent electrical characteristics have been formed using the above process.

Junction depth and impurity surface concentration are important in the stress response of diode sensors. A shallow junction ($\sim$ 1/2 micron) with a high surface concentration ($\sim 10^{21}$ atoms/cm^3) has proved to result in the best sensors. High surface concentration not only increase the stress sensitivity of the sensors but it also makes it easier to obtain good ohmic contact to the needle tip with ordinary metals such as Au and Al.

The stress sensitivity of single junction needle sensors depends heavily on the radius of curvature of the needle point and on the junction area. Since the radius of curvature of the needle points are controllable above a few microns, it is possible to choose almost any desired sensitivity. Typical sensors fabricated in the laboratory have used a radius of curvature of approximately 1/2 mil which resulted in sensors requiring a force of approximately 10 grams to bias the device. The electrical characteristics of a typical sensor is shown in Fig. 2-8 for several stress levels.

The feasibility of fabricating multijunctions on the apex of silicon needles has been investigated. The particular device chosen for this study was a four-layer (three junction) diode. The break-over voltage of four-layer switches is stress dependent (see Appendix A) which makes the device particularly attractive for use in an oscillator configuration. The method studied for the fabrication of four-layer switch needles was identical to that used in fabricating single
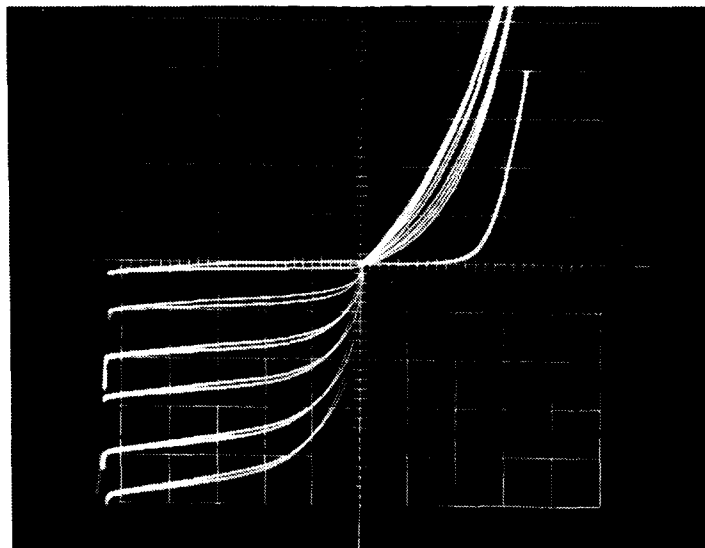
Fig. 2-8. Needle Diode Current-Voltage Characteristics for Several
Stress Levels. Vertical Scale - 0.01 ma/div, Horizontal -
0.2 V/div for Forward Bias and 5 V/div for Reverse Bias.

junction devices as described above except three consecutive dif-
fusions were performed instead of one.

The problem of fabricating four-layer switches in which all three
junctions are diffused junctions is difficult when all the planar pro-
cessing steps are available. That is, when a new oxide can be grown
and a new hole (smaller area) cut in the oxide after each diffusion
step as is common in the planar process. Epitaxial substrates are
much easier to work with in making multijunction devices. Aside from
the masking problem in all diffused junction four-layer diodes each
consecutive diffusion must be much heavier than the previous one at
the expense of controlled junction depth and the quality of the junc-
tion formed.

Since it is not practically feasible using the needle structure to cut more than one window in the oxide on the apex of the needle, it is therefore necessary to carry out all diffusion through a single oxide window.

Considerable effort was put into fabricating four-layer diodes on silicon wafers utilizing a single window for the diffusion. The final process consisted of first making a very shallow (less than 1/2 micron) diffusion with a low surface concentration. This was followed by a drive-in cycle in a neutral furnace, at 1300°C, for approximately six hours. This resulted in a junction depth of approximately 10 to 15 microns. This was followed by two more diffusions with opposite type impurities. The yield on these devices was very low. The problem was not so much control over the diffusions as it was shorted junctions at the oxide junction interface. Another problem was large variations in the device properties, even for devices made at the same time. An analysis of these results showed that these variations resulted in large part from surface effects.

Attempts to translate the single window planar process for four-layer diodes to the silicon needle were unsuccessful. Although a few needle four-layer diodes were formed, they had very poor electrical characteristics. The results indicated, however, that such a process is possible if perfected and controlled properly. It is felt that a near pilot line operation would be required to perform an adequate evaluation of the process. Since neither the funds nor the equipment were available under the present study, further studies were not made.

2.4 Stress and Displacement of Needles Under an Applied Force

The following analysis of silicon needles under loaded conditions is based on needles with circular symmetry, i.e., the shank is assumed to be round. Figure 2-9 shows a sketch of the geometry considered. Silicon is shown in the figure as the pressure plate (base) onto which the needle tip is pressed. Stainless steel, silicon and quartz have been used as base materials.

If the applied force is F, then the displacement of the shank, $\Delta \ell_1$, is

$$\Delta \ell_1 = \frac{F \ell_1}{\pi a^2 E} , \qquad (2.1)$$

where E is Young's modulus and a is the radius of the shank. The displacement of the conical section, $\Delta \ell_2$ is

$$\Delta \ell_2 = \frac{F \ell_2}{\pi a R E} , \qquad (2.2)$$

where R is the radius of curvature of the needle point.

To calculate the displacement of the point, it is assumed that the base dimensions are much greater than the radius of the point. The displacement of the point and the base combined, Δy, is approximately[1]

$$\Delta y \simeq 1.5 \sqrt[3]{\frac{F^2}{E^2 R}} \qquad (2.3)$$

The total displacement of the needle under an applied force F is then

$$\Delta \ell \simeq \frac{F \ell_1}{\pi a^2 E} + \frac{F \ell_2}{\pi a R E} + 1.5 \sqrt[3]{\frac{F^2}{E^2 R}} . \qquad (2.4)$$
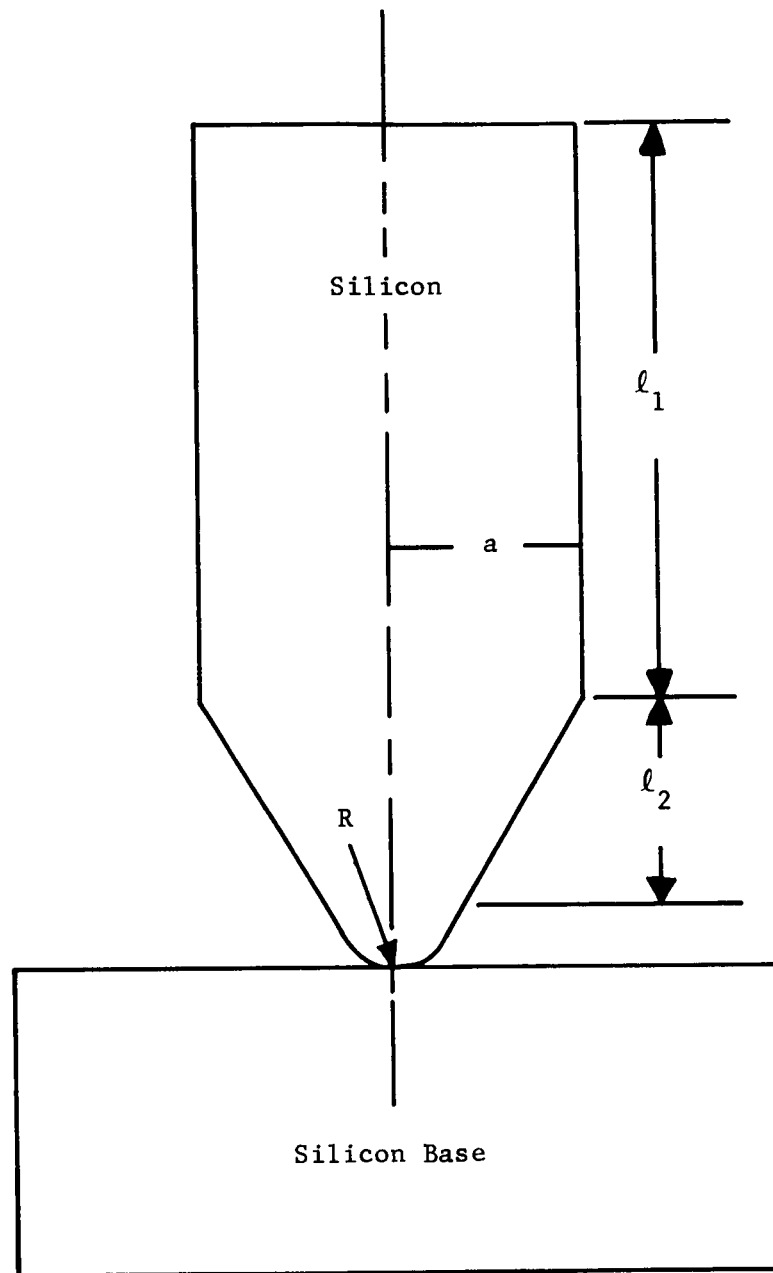
Fig. 2-9. Needle Configuration Used for Calculating Displacement.

As an example, consider a needle with the following typical dimensions:

$$R = 1 \text{ mil}$$

$$\ell_2 = 1/8 \text{ in.}$$

$$\ell_1 = 1/4 \text{ in.}$$

$$E = 1.7 \times 10^{12} \text{ dynes/cm}^2 .$$

In this case:

$$\Delta y \simeq 8 \times 10^{-6} F^{2/3} \text{cm, (F in grams)}$$

$$\Delta \ell_1 \simeq 0.17 \times 10^{-6} F$$

$$\Delta \ell_2 \simeq 0.9 \times 10^{-6} F$$

As can be seen, the total displacement per gram load is approximately one tenth of a micron which is very small. In fact, for most applications, the displacement of the needle can be neglected in comparison to the displacement of the housing or mechanical configuration.

The relationship between stress in the needle tip and applied force is very complicated. This problem has been considered by a number of workers.[2-5] In order to calculate the effects of stress on junction properties, it is necessary to know the stress as a function of position in the needle tip. This is a formidable problem, the solution of which is not at present worth the effort required to obtain it. There has been some success with the evaluation of an indenter point in contact with a planar junction structure.[2]

Since the function $\gamma_v (\sigma)$, which is of particular interest here, is an exponential function of stress, the maximum stress is important. The pressure on the surface under the contact is[5]

24

$$\sigma = \frac{3F}{2\pi a^2} \sqrt{\frac{a^2 - r^2}{a^2}} \, , \qquad (2.5)$$

where a is the radius of the contact circle and r is the distance from the center to the point of interest. The radius of contact, a, is[5]

$$a = \sqrt[3]{(\frac{1 - \nu^2}{E})\,(3/2)RF} \, , \qquad (2.6)$$

or

$$a = K(RF)^{1/3} \, , \qquad (2.7)$$

where ν is Poisson's ratio. The maximum stress is then

$$\sigma_{max} = \frac{3F}{2\pi a^2} \, , \qquad (2.8)$$

or

$$\sigma_{max} = K_2 \, F^{1/3} \, . \qquad (2.9)$$

At high stress levels such that $\gamma_\nu(\sigma) = C_1 \, \exp(C_2\sigma)$

$$\gamma_\nu(\sigma) \simeq C_3 \, \exp(C_4 F^{1/3}) \, . \qquad (2.10)$$

For very shallow junction depths such that the stress is not very different from that at the surface, one would expect the above formulas to be reasonably accurate.

The average stress on the needle tip is

$$\sigma_{ave} = \frac{F}{\pi a^2} \, . \qquad (2.11)$$

Note that the average is a factor of 2/3 smaller than the maximum. The effective value needed in making computations is somewhere between the average and the maximum.

2.5 Electrical Response of Silicon Needle Diodes Under Stress

The single junction silicon needle has been the most used sensor
in this study. The following discussion considers, theoretically, the
electrical response of such devices under stress. The forward and re-
verse biased current-voltage characteristics of diodes under stress are
considered in I and in Appendix C. Both the "ideal" and generation-
recombination currents are considered. The total current as a function
of voltage and stress (stress is represented by $\gamma_\nu(\sigma)$) for a uniformly
stressed junction is

forward bias:

$$I_T = I_{oI}\gamma_\nu(\sigma)\,e^{qV/kT} + I_{or}\sqrt{\gamma_\nu(\sigma)}\;e^{qV/2kT} \qquad (2.12)$$

reverse bias:

$$I_T = I_{oI}\gamma_\nu(\sigma) + I'_{or}\sqrt{\gamma_\nu(\sigma)}\;(V + V_o)^{1/2} \qquad (2.13)$$

Generally speaking, the most promising mode of operation for a
diode needle sensor is the constant voltage mode in which the diode
current is used as the stress indicator. The current is exponentially
related to stress through $\gamma_\nu(\sigma)$ while the voltage is linearly related
to stress. The current is, therefore, more sensitive than voltage.
For the present purposes, it is assumed that voltage is held constant.
Also, since the sensors used were with a dc bias force, the approxi-
mation

$$\gamma_\nu(\sigma) = C_1\,e^{C_2\sigma}$$

is a good assumption. The current can be written as follows

$$I_T = K_1 \, e^{C_2\sigma} + K_2 \, e^{C_2\sigma/2} \, , \tag{2.14}$$

where K_1 and K_2 are constants that depend on the magnitude and polarity of the voltage, unstressed current, crystal orientation and temperature. Normally, K_1 will be much larger than K_2 in the forward biased condition while K_2 will be much larger than K_1 in the reverse biased state. Since both terms in Eq. (2.14) have the same functional relationship on σ, only one term will be considered further, i.e.,

$$I = K_3 \, e^{C_1\sigma} \, . \tag{2.15}$$

As can be seen from Eqs. (2.9) and (2.11), σ is expected to be proportional to $F^{1/3}$. If the total junction area is stressed, then the diode current is functionally related to force as follows:

$$I = K_3 \, e^{CF^{1/3}} \, . \tag{2.16}$$

In most practical sensors, it is more likely that only a part of the junction is stressed. In this case, if A is the total junction area and A_S is the stressed area, the current is

$$I = \frac{A - A_S}{A} \, K_4 + \frac{A_S}{A_T} \, K_5 \, e^{CF^{1/3}} \tag{2.17}$$

The stressed area is related to the radius of contact by

$$A_S = \pi a^2 \, . \tag{2.18}$$

The contact radius is proportional to $F^{1/3}$ as shown in Eq. (2.7). Therefore, $A_S \propto F^{2/3}$. The current is then related functionally to the applied force as follows:

$$I = K_4 - K_6 F^{2/3} + K_7 F^{2/3} \, e^{CF^{1/3}} . \tag{2.19}$$

If $A \gg A_S$, Eq. (2.19) reduces to

$$I = K_1 + K_7 F^{2/3} \, e^{CF^{1/3}} . \tag{2.20}$$

In utilizing the needle sensor in practical transducers, it will be necessary to calibrate the current-voltage characteristics as a function of applied force. Although the functional relationships aid in the design of devices and an understanding of their behavior, the approximations made in the theory are not good enough and the models on which they are based are not accurate enough for device use. This does not mean, however, that the needle sensor is no good. Like almost all of the conventional transducer sensors that are commercially available, they must be calibrated individually.

Section III

ACCELEROMETER DEVELOPMENT

3.1   Introduction and Discussion

One of the major goals of this study has been to design, fabricate, and demonstrate a laboratory accelerometer based on the piezojunction effect. Emphasis has been placed on the silicon needle sensor as the transducing element in the accelerometers. Although a variety of configurations have been studied, they have all been of the diaphragm or beam type. The configuration found to be the best is the double-diaphragm type.

The general case of an accelerometer based on the silicon needle sensor is shown schematically in Fig. 3-1. A dc force biasing spring
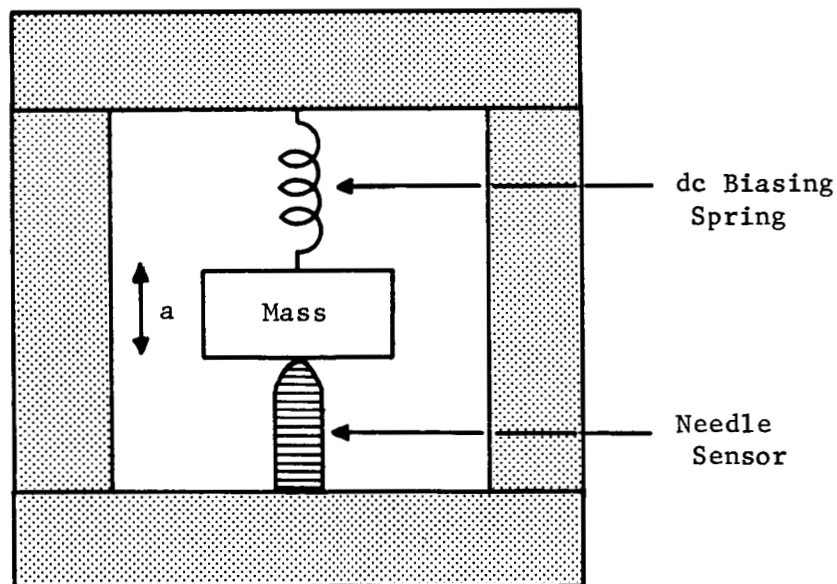
Fig. 3-1.   Schematic of a General Accelerometer Utilizing the Needle Sensor.

(29)

is shown in the figure. This spring can be anything from an actual spring to the give of the mechanical housing holding the mass in place under the bias force applied to the needle.

As discussed in Section 2.4, the displacement of the needle under an applied force F is

$$\Delta\ell \simeq \frac{F\ell_1}{\pi a^2 E} + \frac{F\ell_2}{\pi a R E} + 1.5\sqrt[3]{\frac{F^2}{E^2 R}} \qquad (2.4)$$

This displacement of the needle is important in the dynamic operation of an accelerometer which utilizes the needle sensor. Although $\Delta\ell$ is not linearly related to F, it is approximately linear for small changes in F. It will be assumed for the present purposes that

$$F \simeq K_e \ell \qquad (3.1)$$

where K_e is the equivalent spring constant of the needle which can be calculated from Eq. (2.4).

The dynamic configuration for the accelerometer of Fig. 3-1 is shown in Fig. 3-2. As shown in the figure, K_s is the effective spring constant of the biasing spring and the mechanical housing and M_s is the effective mass of the spring. The resonant frequencies of the system are

$$f = \frac{1}{2\pi}\left\{\frac{1}{2}\left[\left(\frac{K_e}{M} + \frac{2K_s}{M} + \frac{4K_s}{M_s}\right) \pm \left(\frac{K_e^2}{M^2} + \frac{4K_s K_e}{M^2} - \frac{8K_s K_e}{MM_s} + \frac{4K_s^2}{M^2} + \frac{16K^2}{M_s^2}\right)^{\frac{1}{2}}\right]\right\}^{\frac{1}{2}}. \qquad (3.2)$$
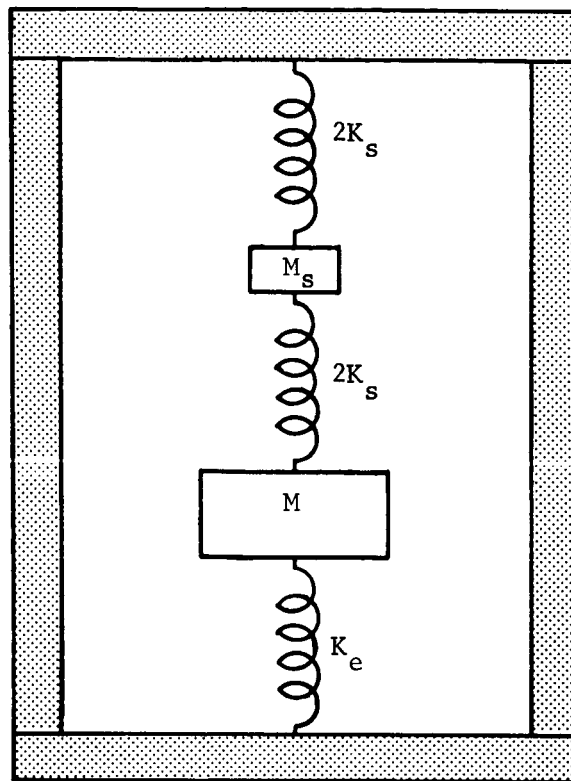
Fig. 3-2. Mechanical Equivalent of an Accelerometer.

If $M_s \ll M$, the fundamental frequency of Eq. (3.2) reduces to

$$f = \frac{1}{2\pi} \sqrt{\frac{K_s + K_e}{M}} \quad . \tag{3.3}$$

The assumption that $M_s \ll M$ is good for low and medium g accelerometers. Also, K_s will normally be much less than K_e.

Using the example of Section 2.4 for the needle sensor and a two gram seismic mass the lowest possible resonant frequency ($K_s = 0$) is approximately 3.7×10^3 cps. An accelerometer utilizing this needle

32

and the two gram weight would, therefore, have a minimum resonant frequency of 3.7 kc.

In the static case where there is no mechanical advantage or disadvantage obtained from the device, the force on the needle is simply,

$$F = Mg + F_o \qquad (3.4)$$

where g is the acceleration level and F_o is the dc bias force.

3.2 Single-Diaphragm Accelerometers

The single-diaphragm accelerometer consists of a cylindrical seismic mass supported by a circular diaphragm with a needle sensor in contact with the diaphragm on the side directly opposite the mass. The diaphragm is clamped or fixed at the edges. Figure 3-3 is a sketch of the accelerometer.

The deflection of the center of the diaphragm is

$$\Delta y = \frac{3(Mg - F)(m^2 - 1)}{4\pi m^2 E t^3} [a^2 - b^2 - (\frac{4a^2 b^2}{a^2 - b^2}) (\log \frac{a}{b})^2] , \qquad (3.5)$$

where Δy = displacement of the mass

   M = seismic mass

   m = reciprocal of Poisson's ratio

   E = Young's modulus of diaphragm

   t = diaphragm thickness

   g = acceleration in g's.

Equation (3.5) can be used to calculate the spring constant of the diaphragm.
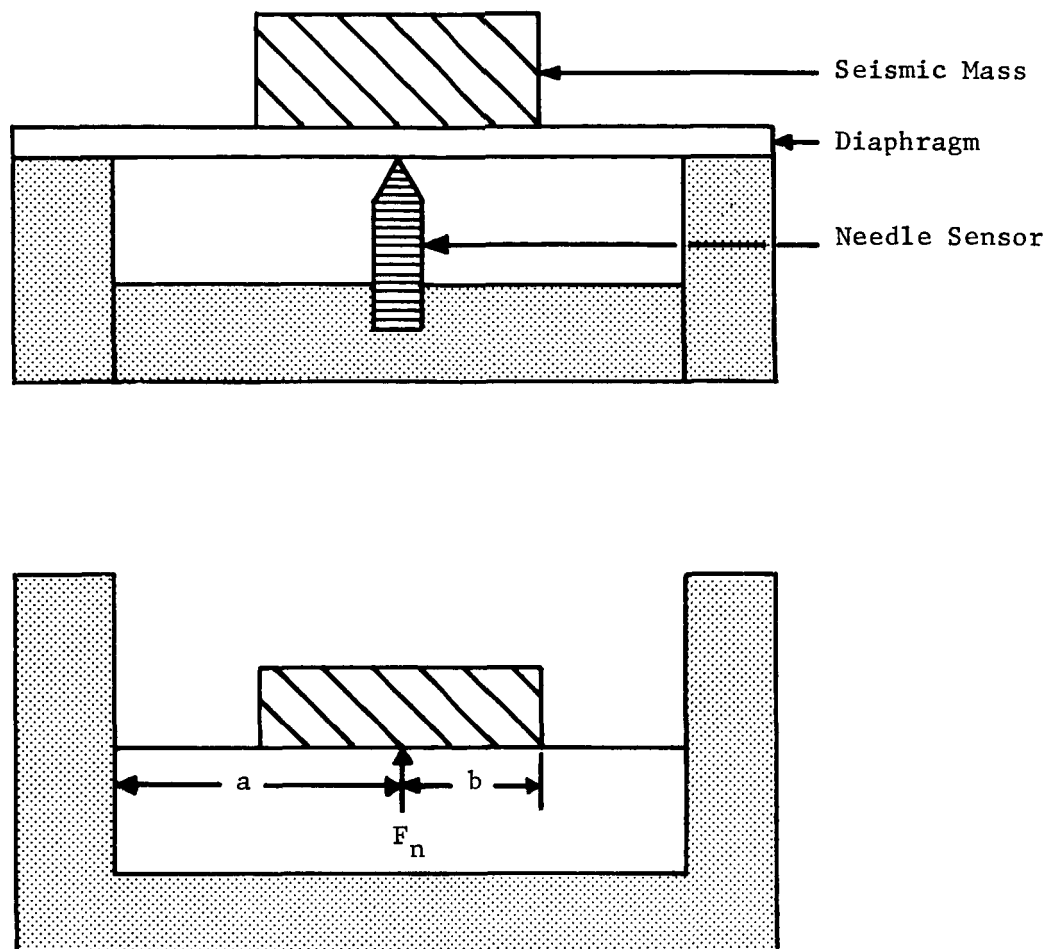
Fig. 3-3. Sketch of Single-Diaphragm Accelerometer.

The dc bias force can be applied in the diaphragm accelerometer by simply constructing the device so that the needle is exerting the desired force against the diaphragm. The diaphragm acts in this case as the biasing spring. Figure 3-4 is a photograph of a single-diaphragm accelerometer in which the diaphragm is used as the biasing spring. The device consists of a brass base in the center of which a silicon needle sensor is soldered, a glass housing, a brass diaphragm (2 mils thick), and the seismic mass. The device is fabricated by first
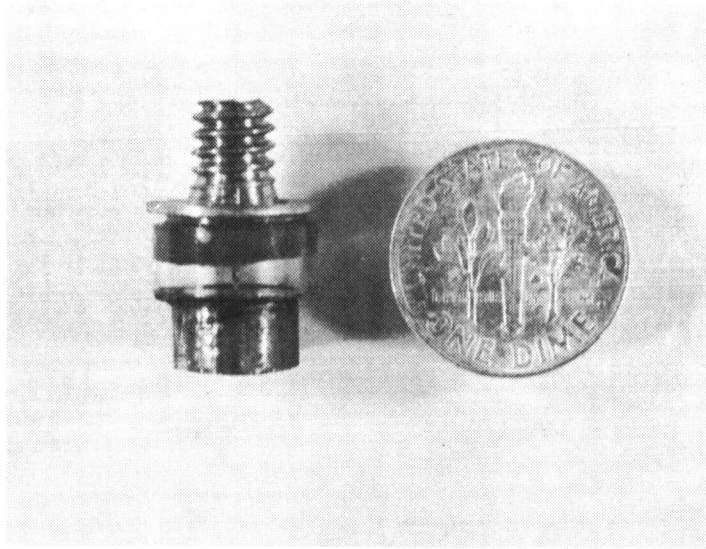
Fig. 3-4. Photograph of a Single-Diaphragm Accelerometer.

soldering the mass to the diaphragm. The diaphragm is then epoxied to
the glass housing. The next step is to attach the needle sensor to the
base. This is done by drilling a hole the size of the needle and
soldering the needle shank into the base. The final step consists of
epoxying the base to the housing. During the curing cycle for the
final step, a constant force is maintained between the glass housing
and the base to create the desired dc bias force. Electrical contact
is made between the diaphragm and the base.

The needle diode sensors used in accelerometers of this type were
pressed into the brass diaphragm. Subsequent tests of these accelerom-
eters showed large hysteresis and drift effects which were eventually

traced to the plastic deformation of the brass under the needle point.
Each time the devices were accelerated to a new high acceleration level
the devices exhibited a new set of electrical characteristics as a
function of acceleration. Since there were no means provided for the
adjustment of the dc bias force, the devices usually ended up by losing
electrical contact to the needle sensor.

In addition to the above mentioned problems, the single-diaphragm
accelerometers were found to be sensitive to cross-axes acceleration.
This was not an unexpected result since a cross-axes acceleration
applies a large moment on the diaphragm due to the mass and height of
the seismic mass.

Some improvements were incorporated into the single-diaphragm
accelerometers. One of the improvements was to use stainless steel
as the diaphragm material. Another improvement was the use of an
adjustable bias force. The latter was accomplished by mounting the
needle on the end of a screw which was then threaded into the base.
Since the needle rotates with respect to the diaphragm during the
assembly operation, it was necessary to pull up on the diaphragm
until the needle was inserted to the desired height and then release
the diaphragm to allow it to make contact with the needle. This pre-
vented breakage of the needle.

Although the single-diaphragm accelerometer is relatively easy
to construct, its cross-axes sensitivity, especially for applications
in the low g range, make it unattractive compared to other devices
such as the double-diaphragm type.

3.3 Double-Diaphragm Accelerometers

The double-diaphragm accelerometer was designed to eliminate the cross-axes sensitivity problems associated with the single-diaphragm type. A schematic drawing of the double-diaphragm type is shown in Fig. 3-5. The second diaphragm prevents the mass from rotating under cross-axes acceleration.

For the case where both diaphragms are identical, the displacement of the mass is given by Eq. (3.5) with (Mg - F) replaced by (Mg - F)/2.
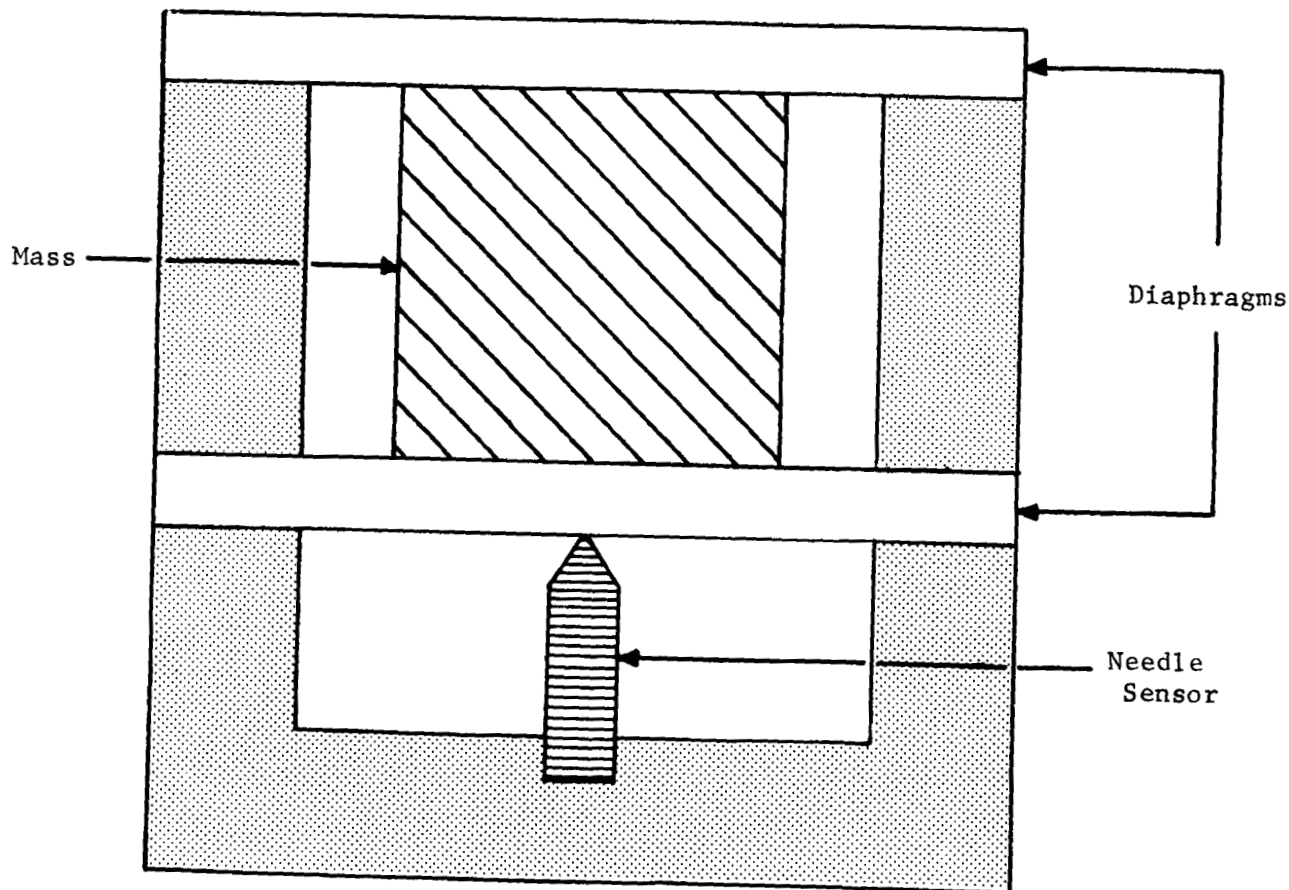
Fig. 3-5. Schematic of Double-Diaphragm Accelerometer.

In cases where the two diaphragms are different, Eq. (3.5) must be applied for each diaphragm with (Mg - F) replaced by (Mg - F - Q) for one and Q for the other. The two equations are then solved simultaneously.

A variety of double-diaphragm accelerometers have been fabricated and tested. Based on the results obtained on the single-diaphragm type and other types, it was decided in the beginning that an adjustable dc bias force was necessary. Since there is always some relaxation of the mechanical housing and some plastic deformation of the needle and its base, the adjustable bias was incorporated into the design. Figure 3-6 shows a schematic representation of the accelerometer showing the

Silicon Needle Sensing Element

Needle (base) Pressure Plate

Seismic weight

Double-Diaphragm Suspension

Spring and Screw Bias Arrangement
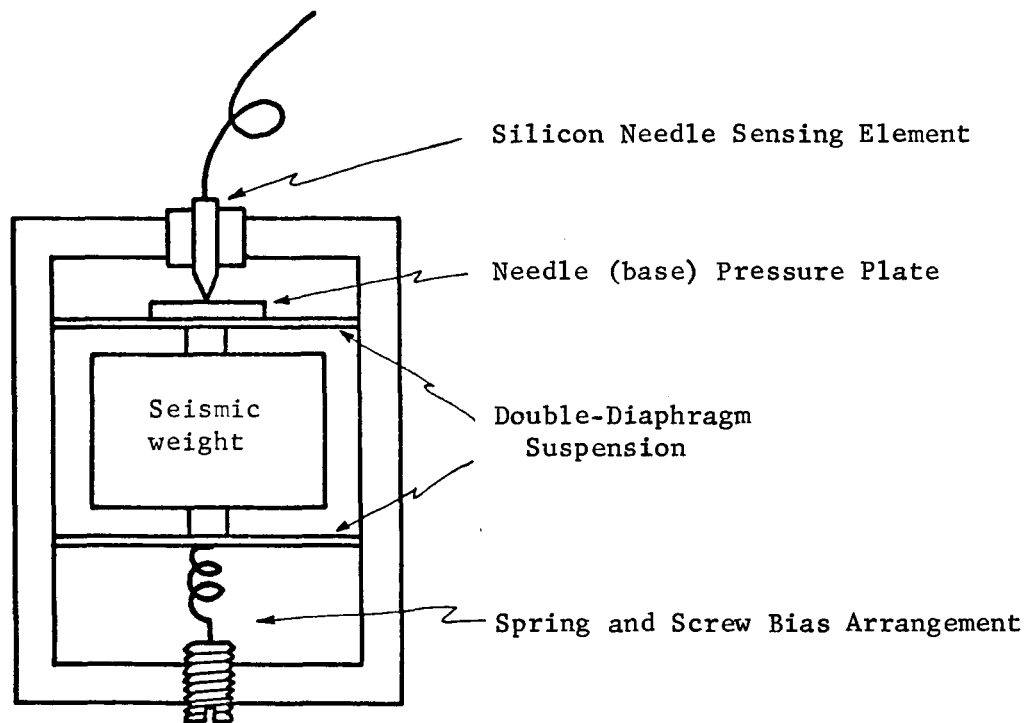
Fig. 3-6. Schematic Representation of a Double-Diaphragm Accelerometer Showing the dc Bias Adjustment Arrangement.

dc bias adjustment arrangement. For this type of arrangement, the equivalent spring constant of the diaphragms and the spring is obtained by simply adding the spring constant of the bias spring, K_b, to the spring constant of the diaphragms, i.e.,

$$K_e = K_d + K_b .$$ (3.6)

It should be noted that the needle is mounted on the opposite side of the diaphragms and mass from the bias spring. If it were mounted behind the needle and the needle allowed to move, the resonant frequency would be lowered considerably and the system would be extremely unstable.

A cross-section view of a typical double-diaphragm accelerometer is shown in Fig. 3-7. The housings for these devices have been brass, with 10-32 screw threads on the bottom for mounting purposes. The housing is electrically ground and the other electrical lead is attached to the needle shank. The shank of the needle sensor is electrically isolated from the accelerometer housing.

The diaphragms are soldered to the center housing ring and the housing is put together with screws. The needle holder is epoxied to the top housing piece. The needle is mounted in a brass rod (holder) by soldering and the rod is then epoxied in the housing. Electrical insulation is provided by the epoxy. In the device shown in Fig. 3-7, the mass is attached to the diaphragms by screws. This can also be accomplished by epoxy or solder.

Laboratory accelerometers were tested both statically and dynamically in the Standards Laboratory at Langley Research Center,
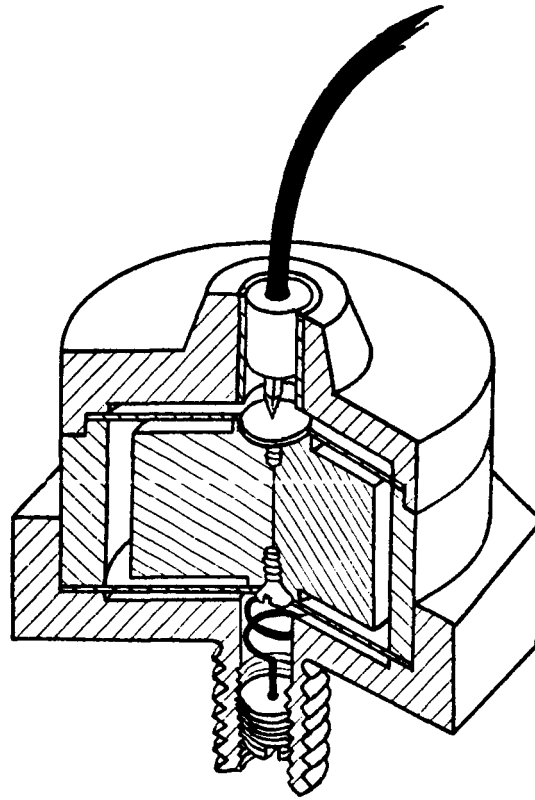
Fig. 3-7. Cross-Section of a Typical Double-Diaphragm Accelerometer.

Hampton, Virginia. In these tests, a positive acceleration was such that the seismic mass pressed against the needle sensor and a negative acceleration relieved stress on the needle.

Three series of accelerometers of the double-diaphragm type have been fabricated and tested. All three series were of the same general configuration as that shown in Fig. 3-7. The first series utilized a gold plated stainless steel pressure plate (base) for the silicon needle. The needle sensors in this set did not have any metal contact material on the tip of the needle, i.e., electrical contact was made by the silicon needle tip in contact with the gold plated steel.

Figure 3-8 is a photograph of one of the accelerometers of the

first series. These devices had a 2 gram mass. The top of the

accelerometer was designed to accommodate a Microdot connector for

electrical contact. These devices were tested by using the current at

a constant voltage as a measure of the acceleration. Figure 3-9 is

a plot of ΔI as a function of acceleration with a reverse bias voltage

of -15 volts. The curve is a copy of the data as recorded on an x-y

plotter. As can be seen in Fig. 3-9, there is a hysteresis in the

characteristics. The current did not return to its original value

following an acceleration to a new high level. When tested at accel-

eration levels below some previous level, the loop was found to close

at zero g. Another example of the hysteresis effect in another device

is shown in Fig. 3-10. The top curve is the first run and the bottom

curve is the second run.

Tests of cross-axes sensitivity showed that the devices had a

cross-axes sensitivity of less than 1% of that in their sensitive axis.

Figure 3-11 is a plot of sensitivity at a constant g level as a function

of frequency. As shown, the resonant frequency was approximately 3 KC.

This is slightly lower than calculated earlier. This was expected

since the pressure plate was stainless steel and not silicon as had

been assumed in the calculations.

Aside from the hysteresis effects in the first set of devices, the

sensitivity was much lower than expected. A careful analysis of the

test results showed that the low sensitivity and much of the hysteresis

effects resulted from plastic deformation of the stainless steel

pressure plate.
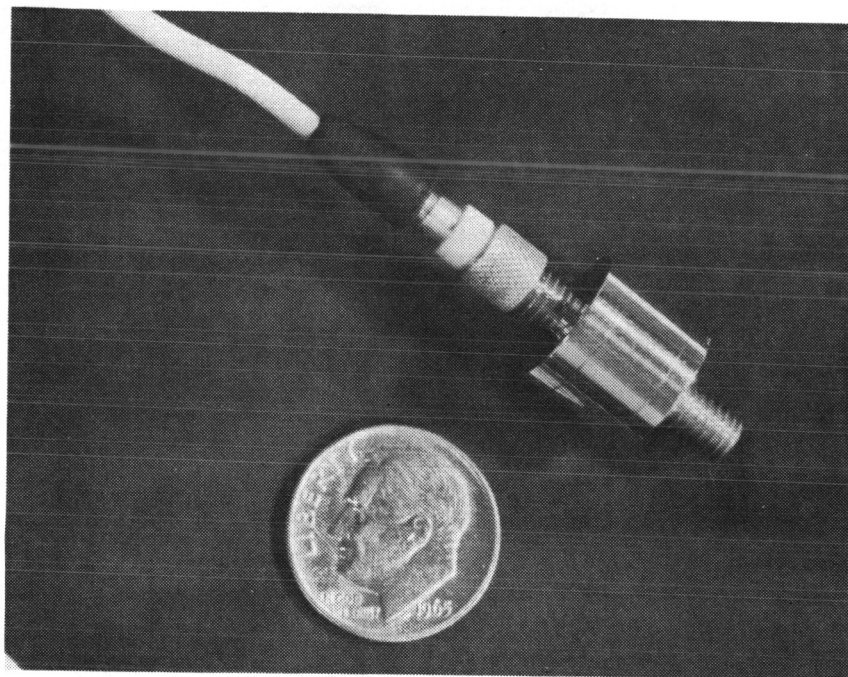
Fig. 3-8. Photograph of a Double-Diaphragm Accelerometer with a
2 gram mass.

42
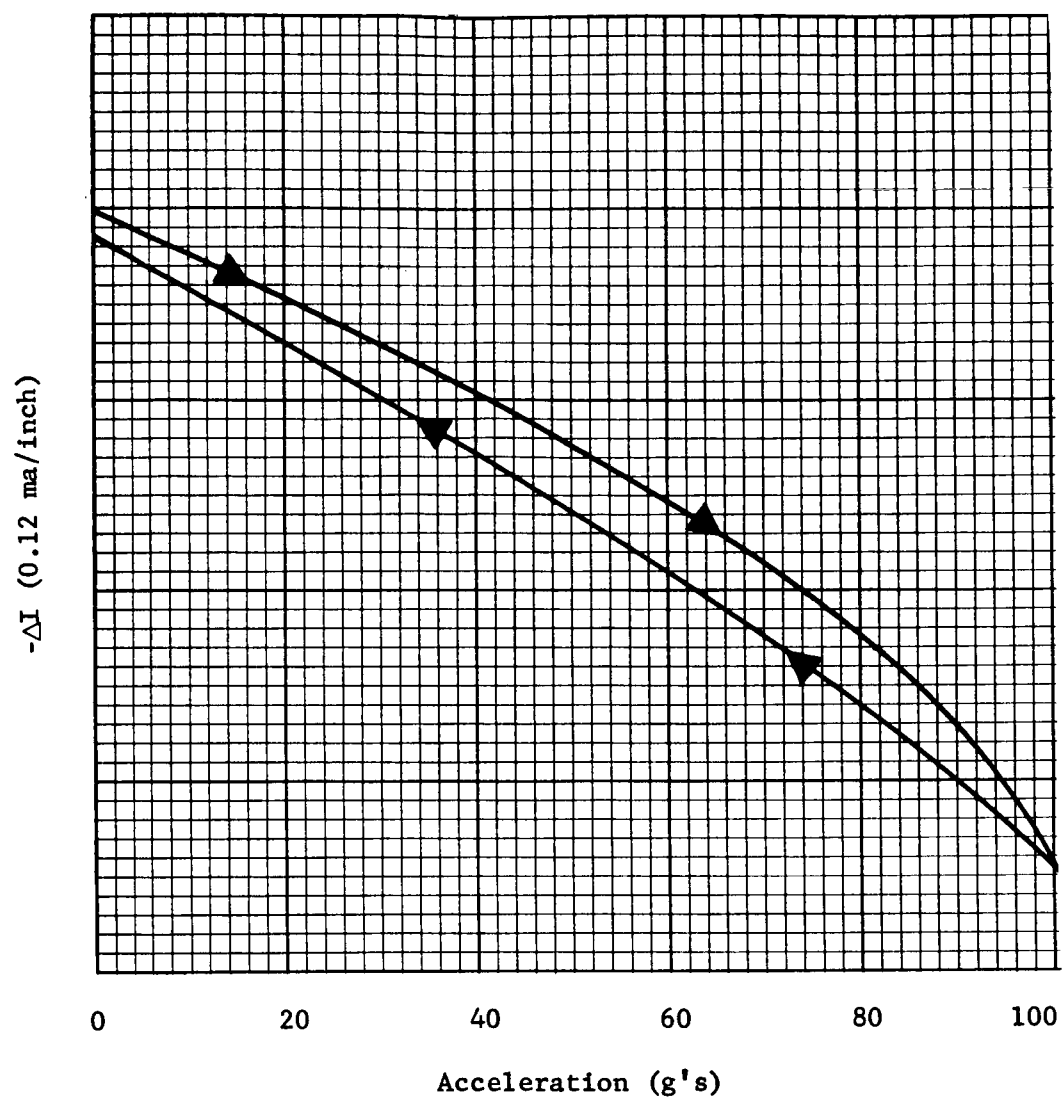
Fig. 3-9.   Change in Current (V = -15 volts) as a Function of
            Acceleration for the Accelerometer Shown in Fig. 3-8.
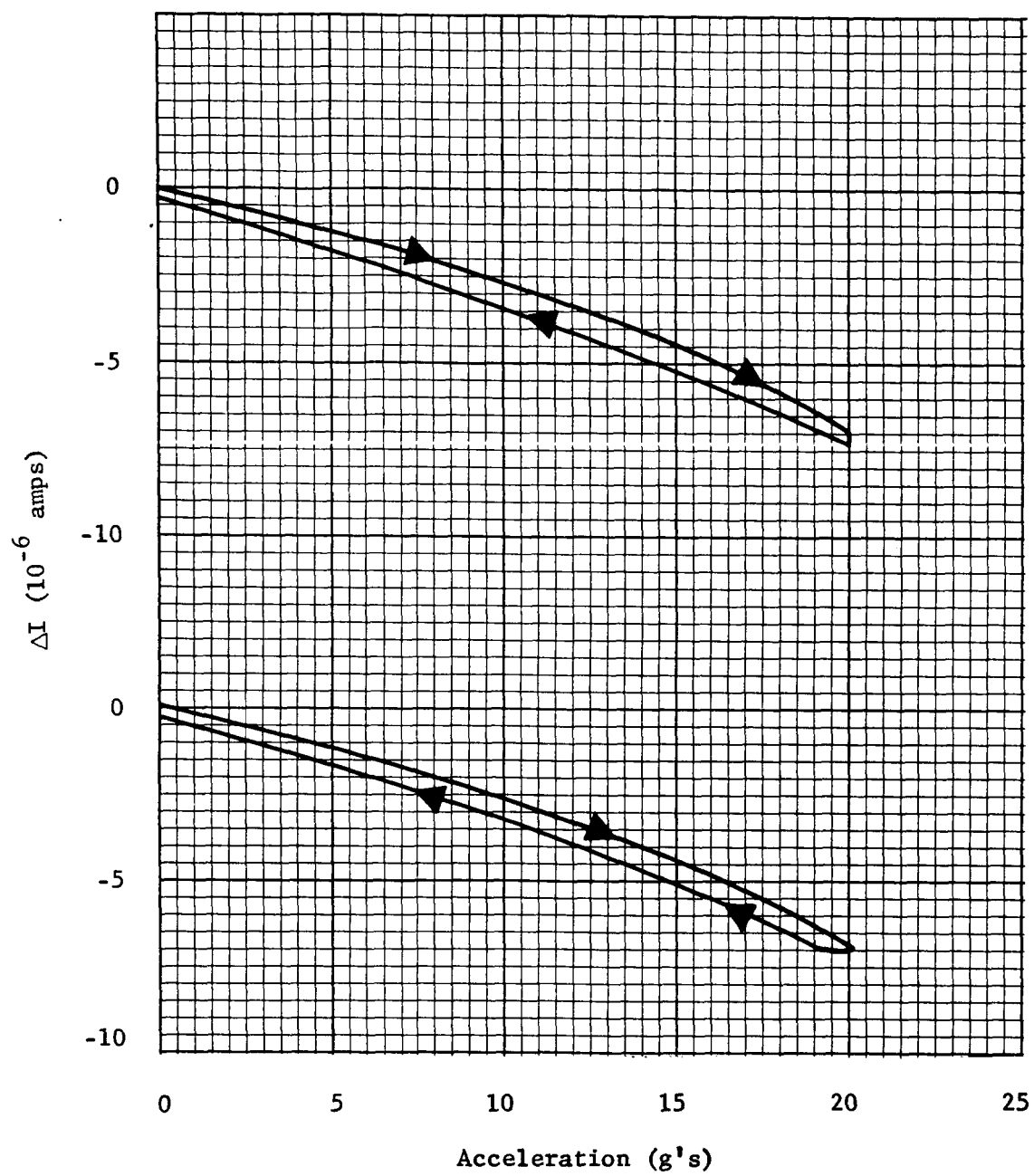
Fig. 3-10. Change in Current as a Function of Acceleration for an
           Accelerometer of the First Series.
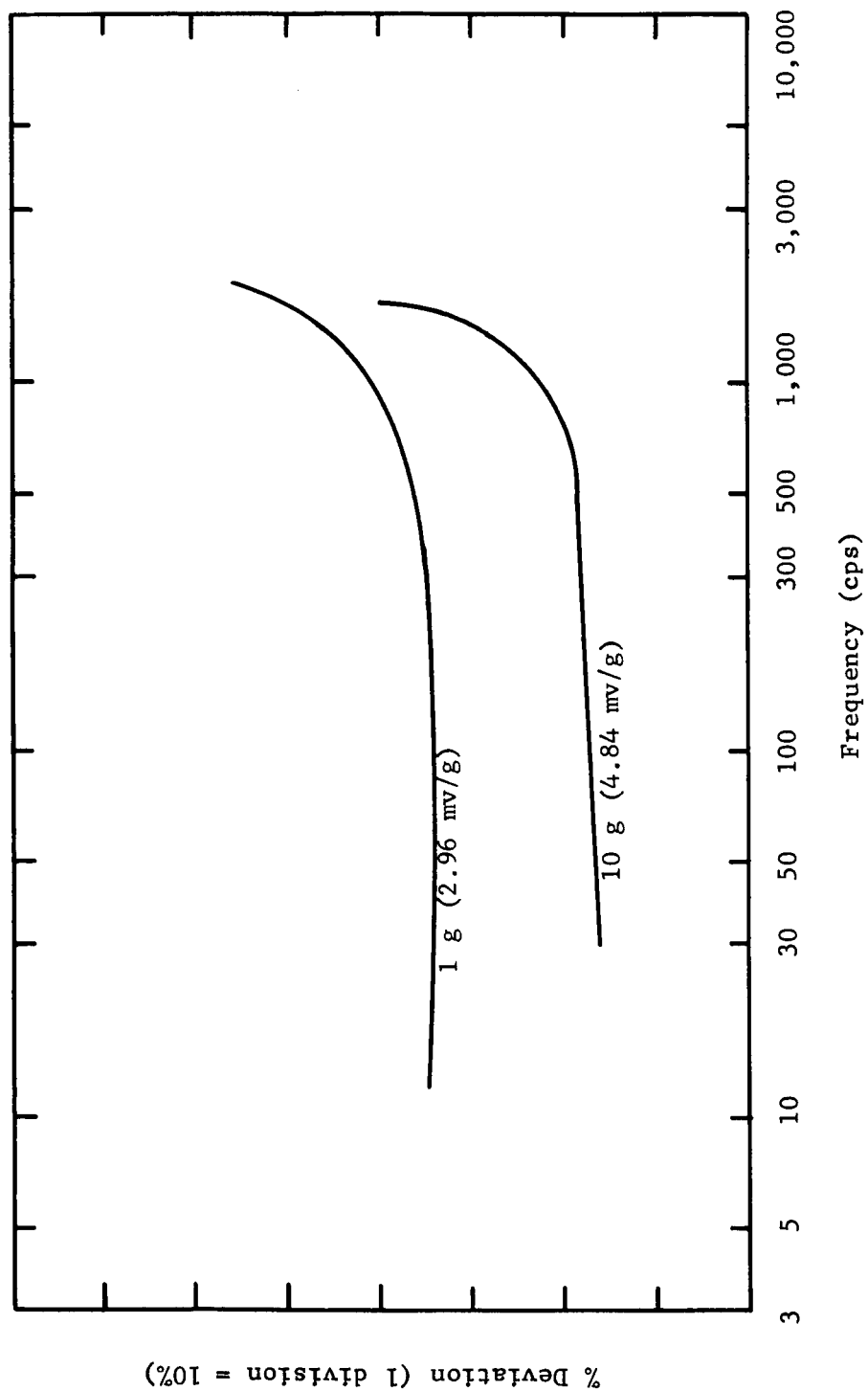
Fig. 3-11. Plot of Sensitivity as a Function of Frequency for 1 g and 10 g Sinusoidal Accelerations.

The second set of devices that were fabricated and tested were identical with the first set except the pressure plate was made of gold plated silicon. This made the mechanical properties of the pressure plate and the sensor identical. Again, these devices did not utilize any metal on the tip of the silicon needle. Although several of these devices were fabricated, all but one was broken while being transported to LRC for testing. The one remaining device was tested with the aid of the circuit shown in Fig. 3-12. The offset or bias voltage V_o is the voltage required in series with the recorder to give a zero voltage under zero g acceleration. V is the battery voltage. ΔV is the change in voltage across the 10 K resistor due to acceleration.

Figure 3-13 is the actual recorder plot of ΔV as a function of acceleration for several reverse biased conditions. As shown, the acceleration was from -5 to +5 g's. The vertical scale for ΔV is 1 volt per major division (1 inch). As shown in Fig. 3-13, both the sensitivity and the hysteresis effect increases with reverse bias voltage.

The forward biased mode is shown in Fig. 3-14. Again, the sensitivity increases with bias voltage. Note in Fig. 3-14 that two curves are shown for V = 0.4 volts. The lower 0.4 volt curve was slightly offset with the recorder zero and represents four consecutive runs to show the repeatability. As can be seen by comparing Fig. 3-13 with 3-14, the hysteresis effect is virtually non-existent in the forward biased case.
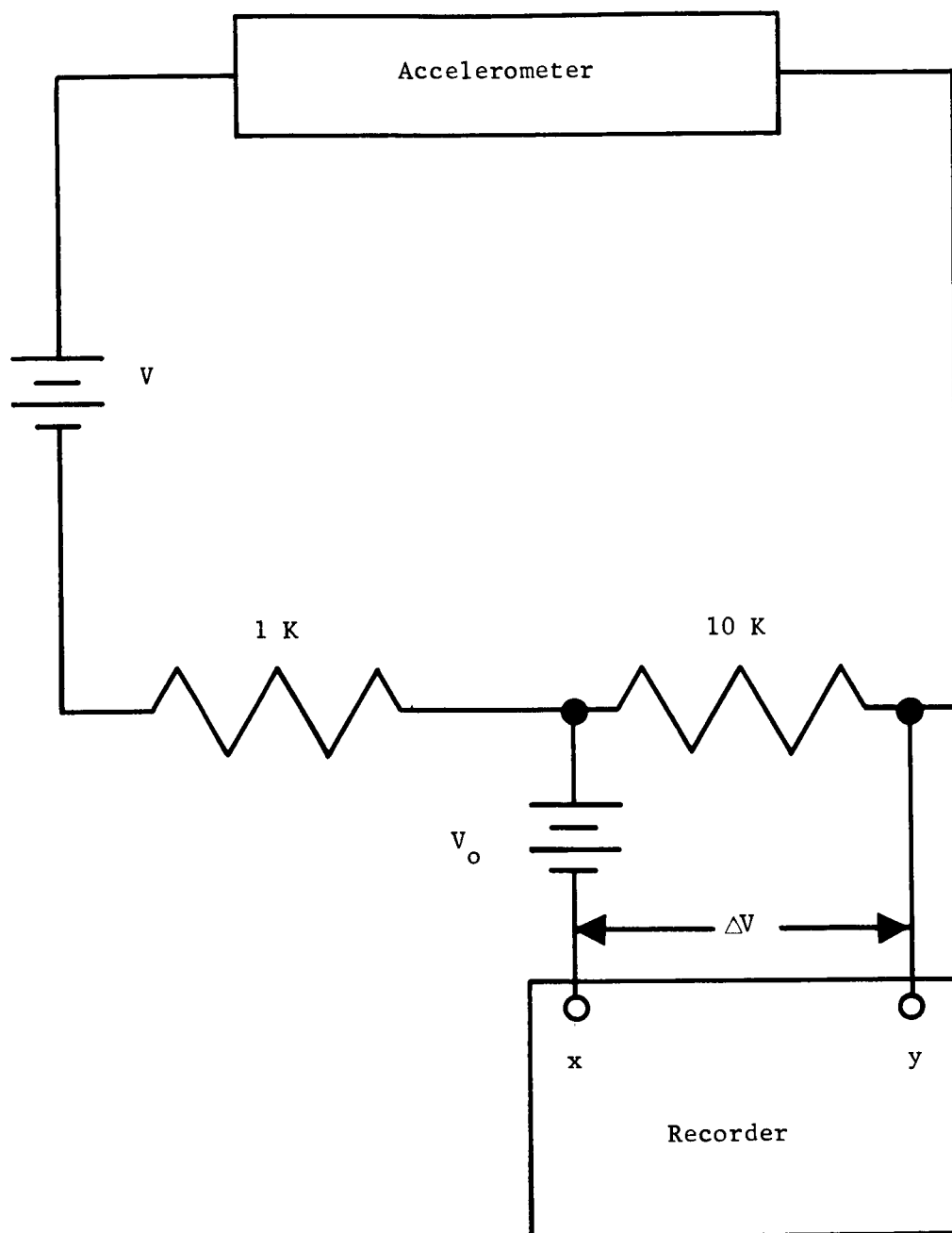
Fig. 3-12. Circuit Used to Test Accelerometers.
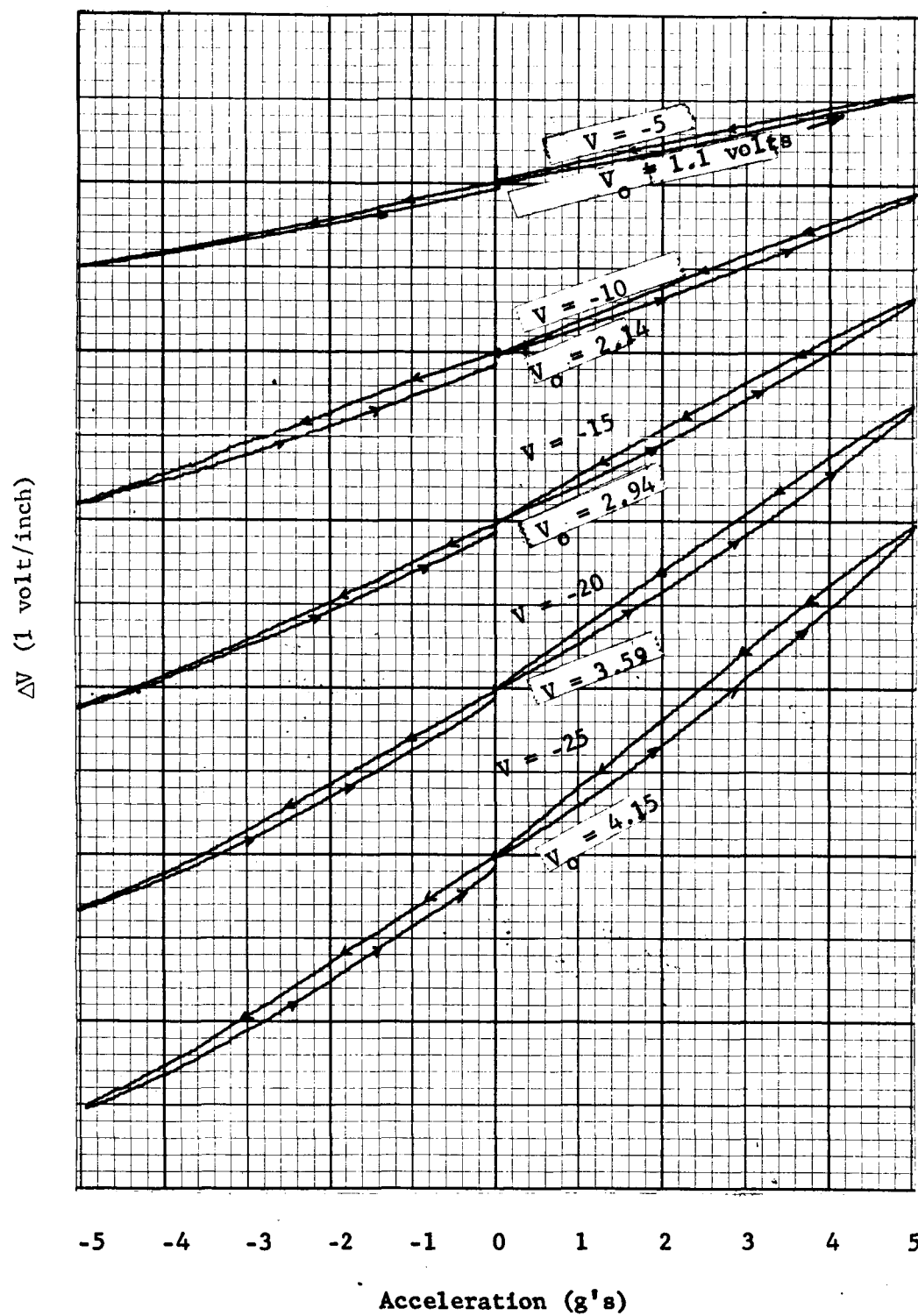
Fig. 3-13. Recorder Plot of ΔV as a Function of Acceleration for the
Reverse Biased Mode. The Vertical Scale is 1 volt/major division.

48
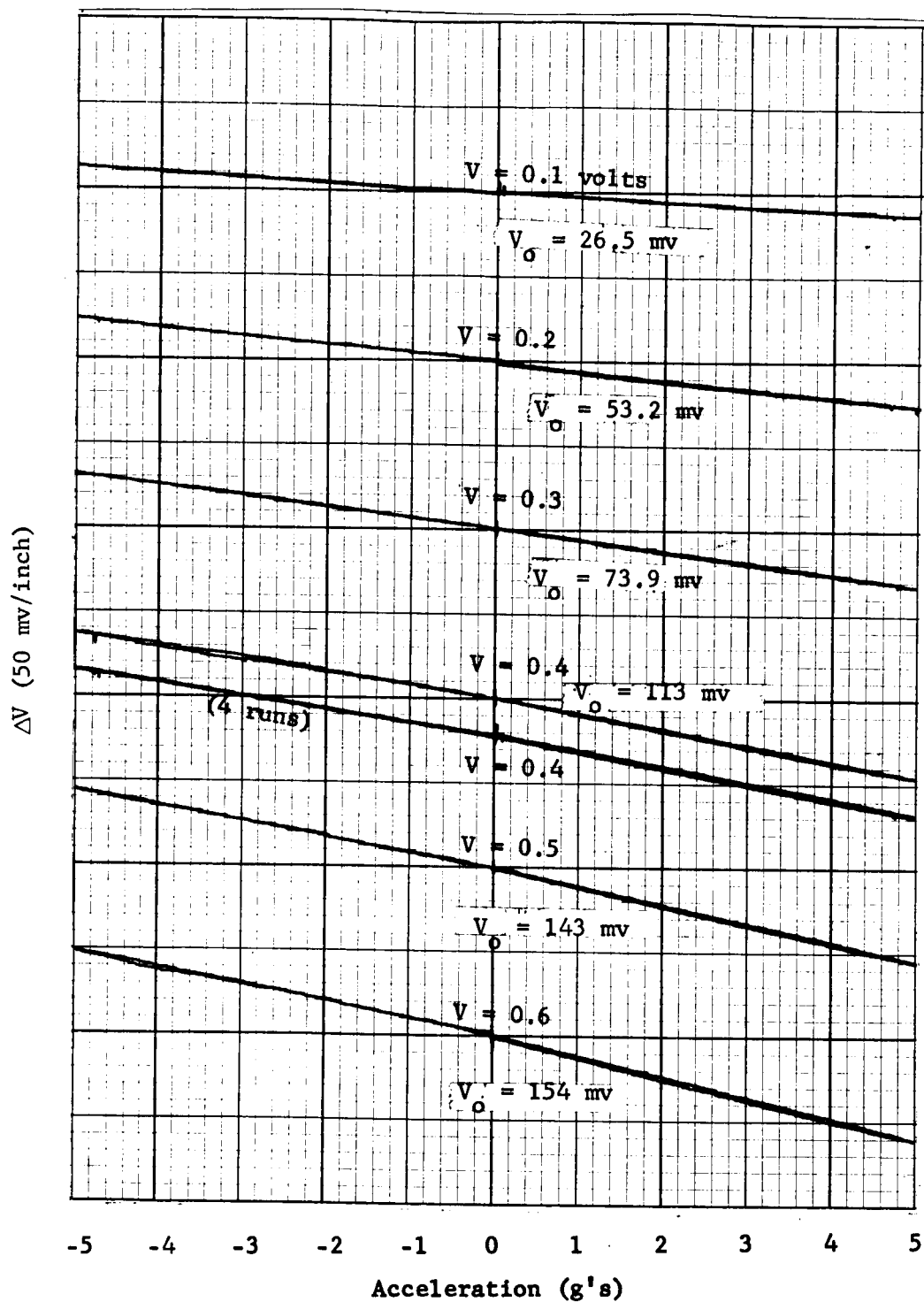
Fig. 3-14. Recorder Plot of ΔV as a Function of Acceleration. The Device was Operated in the Forward Biased Mode. The Vertical Scale is 30 mv/major division.

Figure 3-15 is a recorder plot of sensitivity for a reverse bias of -15 volts for acceleration along the sensitive axes and along the transverse axes.

The third set of devices utilized the gold plated silicon pressure plate. In addition, the needle tip was coated with an uncured insulating epoxy prior to installation of the needle into the housing. The needle point punched through the epoxy when the bias force was applied. The idea here was to add mechanical support to the needle especially for transverse motions. Also, the epoxy should provide some heat sink effects. As one might expect, the devices were plagued with poor electrical contacts. These poor contacts were experienced even though gold was plated on the needle tips prior to the epoxy.

Figure 3-16 is a photograph of one of the accelerometers of the above type. This particular device was designed to respond to low g accelerations by making the radius of curvature of the needle tip smaller than the previous devices. Figure 3-17 is a photograph of the forward and reverse current-voltage characteristics of the accelerometer under 0 and $\pm$ 1 g accelerations -- A = -1 g, B = 0 g, C = + 1 g. The vertical scale is 0.1 ma/division for both the forward and reverse modes. The horizontal side is 0.5 volts/division for the forward and -5 volts/division for the reverse mode. The non-linearity of the piezojunction effect can easily be seen in Fig. 3-17 where, for a constant voltage, the current change between 0 and -1 g is seen to be smaller than the change between 0 and + 1 g.
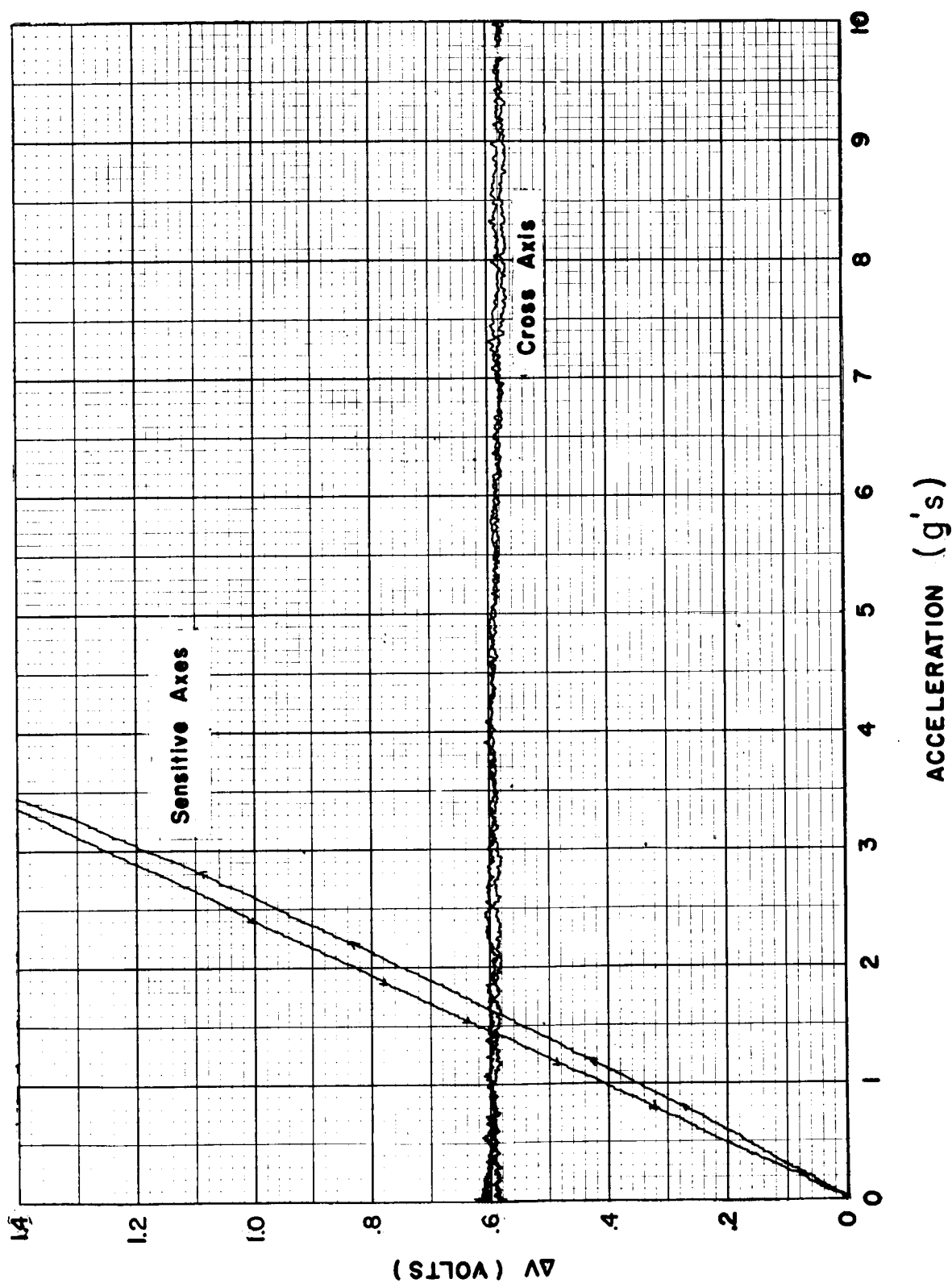
Fig. 3-15. A Recorder Plot of Sensitivity for a Reverse Bias of -15 Volts for Acceleration Along the Sensitive Axes and Along the Transverse Axes.
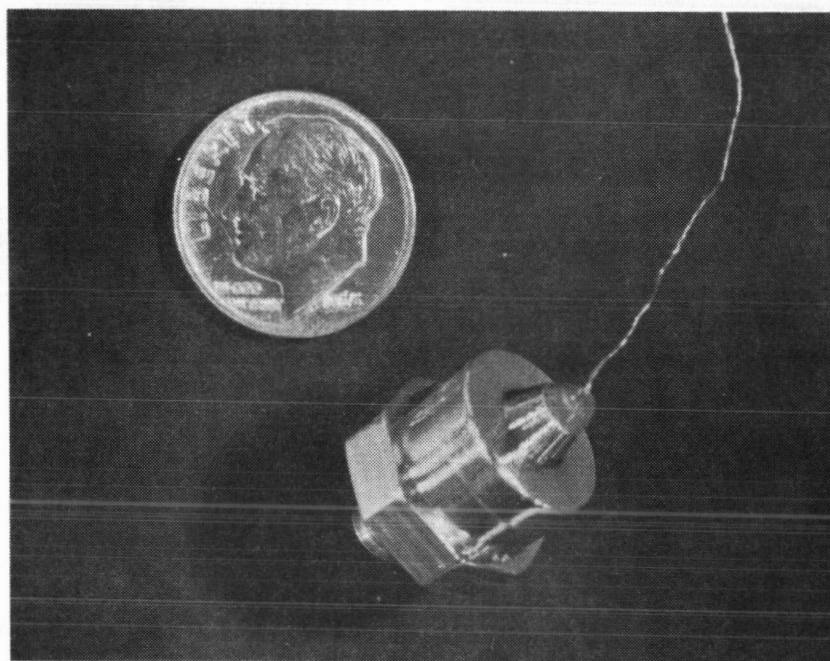
Fig. 3-16. Photograph of a Double-Diaphragm Accelerometer. No
Microdot Connector is Provided.
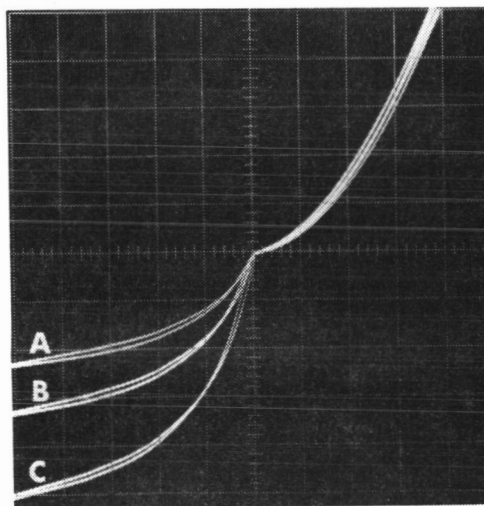
Fig. 3-17. Current-Voltage Characteristics of the + 1 g Accelerometer
Shown in Fig. 3-16. A = -1 g, B = 0 g and C = + 1 g.
The Vertical Scale is 0.1 ma/div. for Both Forward and
Reverse Modes. The Horizontal Scale is 0.5 V/div. for
Forward and -5 V/div. for Reverse Mode.

Figure 3-18 shows a recorder plot of ΔV as a function of acceleration for several reverse biased conditions. The circuit shown in Fig. 3-12 was again used to test this device. Note that the hysteresis effects were also present with these devices and as before was found to increase with an increase in bias voltage.

Figure 3-19 shows a plot of frequency as a function of acceleration for the above accelerometer. The oscillator circuit used here is discussed in Section 5.4. An 18 volt supply was used and a 0.001 μf capacitor. As shown in Fig. 3-19, the sensitivity is approximately 1.4 KC/g.

3.4 Beam Accelerometers

Another possible mechanical configuration for accelerometers which utilize a needle sensor or indentor point junction combination is a cantilever beam arrangement. A detailed discussion of such an arrangement for use as a force and displacement transducer is given in Section 4.3. The mechanical force in the arrangement discussed in Section 4.3 is simply replaced by a mass under acceleration for accelerometer action. The mass can be the mass of the beam, a mass attached to the beam, or both.

Laboratory accelerometers of the cantilever beam type have been investigated. These devices were found to be extremely fragile and easily broken. The resonant frequency of the cantilever beam accelerometers is low (typically less than 1 KC). The major advantages are its simplicity and range. The beam can be used either to give a mechanical advantage or disadvantage and hence a large range of devices
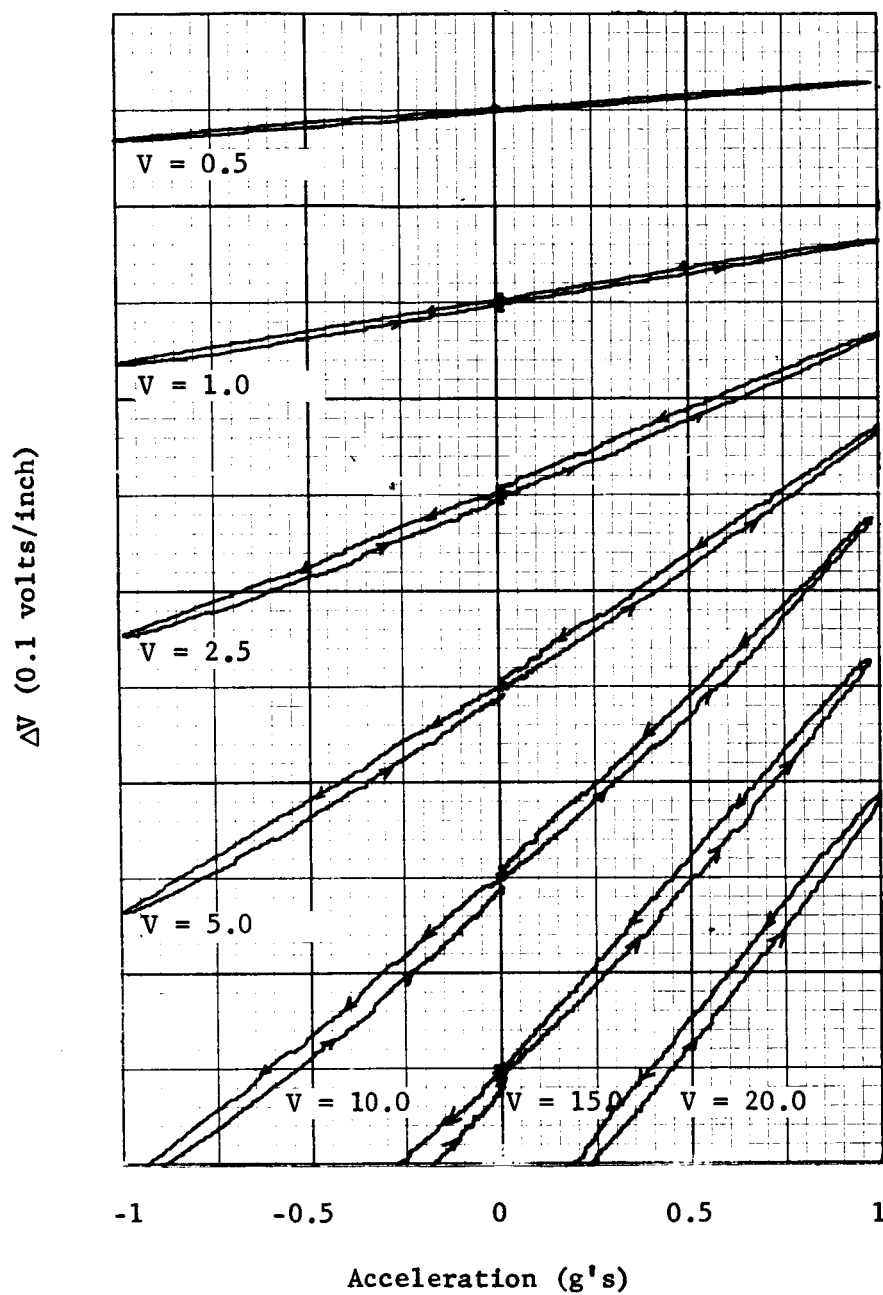
Fig. 3-18. A Recorder Plot of ΔV as a Function of Acceleration for Several Reverse Biased Conditions.
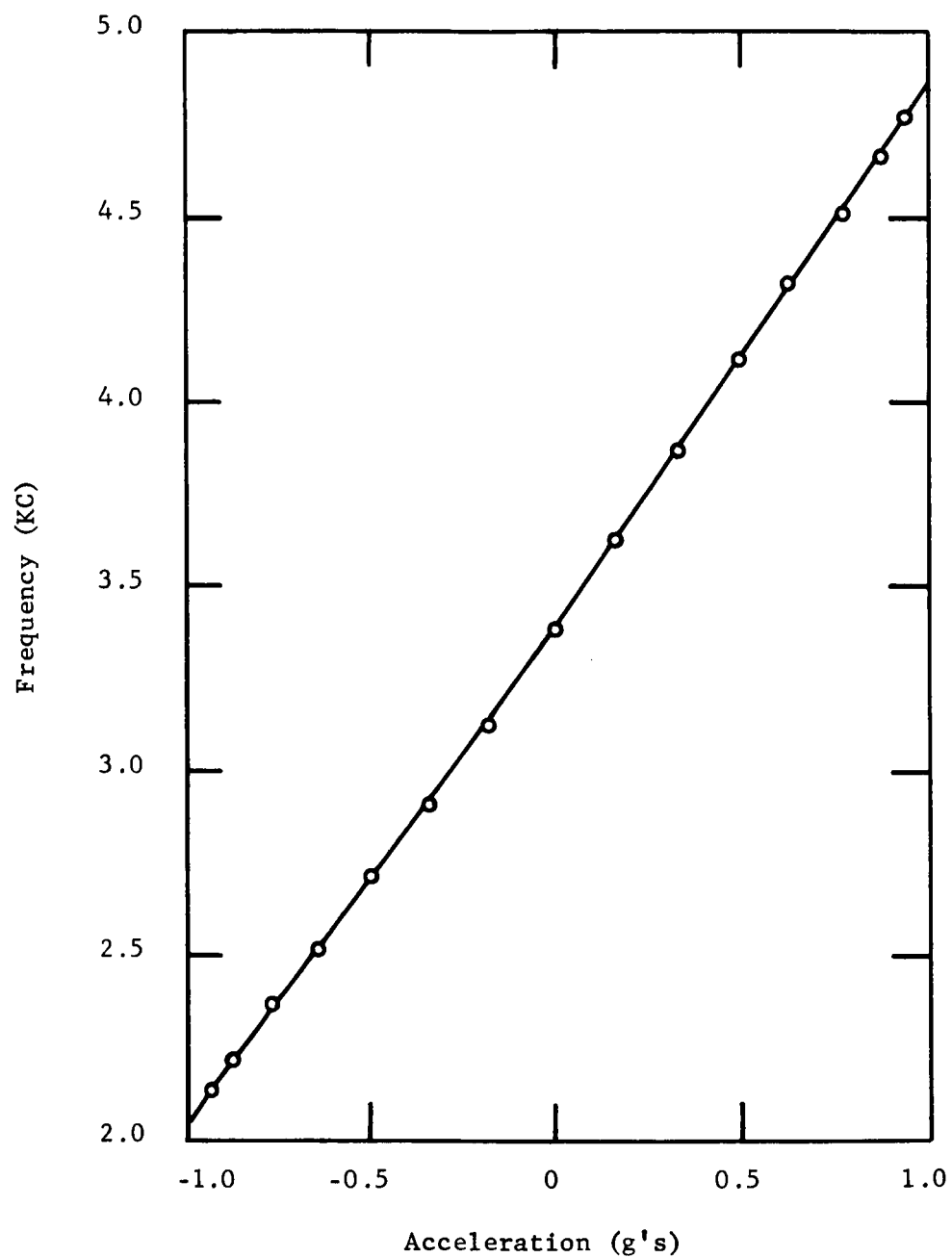
Fig. 3-19. Frequency vs Acceleration for the Accelerometer Shown in Fig. 3-16. The Unijunction Oscillator was Used.

is possible. Low g beam devices are typically sensitive to cross-axis

accelerations, however, this can be kept to a minimum by using wide

beams.

3.5 Discussion and Summary

Accelerometers of the single-diaphragm, double-diaphragm and

beam types have been designed, fabricated and tested. Of the types

studied, the double-diaphragm type is definitely the best suited

configuration for use with silicon needle sensors. The double-diaphragm

accelerometer is, however, the most difficult to construct. As demon-

strated by the data presented herein, accelerometers can be fabricated

in the range $\pm$ 1 g to $\pm$ 100 g. This does not imply that these are the

limiting ranges. In fact, it should be possible to extend both ends

by an order of magnitude by choosing the proper radius of curvature

for the sensor and the proper mass. Resonant frequency can also be

extended by making the bias spring stiffer and reducing the mass both

of which are within practical realization.

The major problem with the devices that were tested was the

hysteresis effects in the output signal as a function of acceleration.

The use of steel pressure plates (bases for the needle) was found to

cause large hysteresis effects and non-reproducibility in the

characteristics. This was not unexpected since the yield strength

of steel is less than the stress levels needed to produce the piezo-

junction effect.

The use of silicon as the pressure plate material was found to

reduce the hysteresis effect and resulted in reproducible characteristics.

An analysis of the data on the accelerometers utilizing the silicon

pressure plates shows that the hysteresis effect is directly related to the sensitivity of the devices. The greater the sensitivity, the greater the hysteresis effect.

Although it is not shown in the data, it was observed that the hysteresis effects were time dependent. The hysteresis loops were found to be smaller when a few minutes were allowed to elapse at each acceleration level. The time required to reach equilibrium was approximately 1 to 2 minutes. As shown by the data, when the acceleration was in the direction of increasing force on the needle, the current was greater on the decending acceleration curve than it was on the ascending portion of the curve. For a negative acceleration, the current was greater when the stress on the needle was being decreased than it was when stress was being increased.

At a first glance, one might suspect that the hysteresis effect is a result of the creation of generation-recombination centers in the junction of the sensor. This model was originally presented as the mechanism responsible for the piezojunction effect. The time dependence is of the correct order of magnitude for such a mechanism. Consider the creation of generation-recombination centers which appear with some very small time constant as stress is applied and which anneal out with a time constant τ after stress is relieved. If τ is much less than the time required to increase and decrease the force on the needle, then there would be no difference observed in the magnitude of the current between an increasing and decreasing force. On the other hand, if τ is equal to or greater than the time required to increase and decrease the force, then a hysteresis effect would be present in

the current-force characteristics. In the latter case, if force were first increased then decreased, the generation-recombination centers introduced during the force increasing period would remain during the decreasing portion of the cycle and would therefore result in a higher current value. Now if force is decreased and then increased, any generation-recombination centers in the material at the time the cycle is initiated would not have time to anneal out before they were re-introduced, the result being no net change in the current. As shown in the acceleration data in Fig. 3-13, the above conclusions are not observed. In the compression mode the loop closes at zero g and there is a hysteresis present in the stress relieving mode. The creation of generation-recombination centers as a mechanism is therefore ruled out as a cause of the hysteresis. If this mechanism was the cause of the hysteresis then it would be a fundamental limitation on the piezojunction phenomenon.

The observed hysteresis effect is believed to be a result of plastic deformation or non-elastic behavior of the pressure plate material. If the plastic deformation were in the needle it would re-sult in permanent changes in the current or at least generation-recombination centers as discussed above would be created. The contact area between the needle and pressure plate and the stress in the needle would both be affected by plastic deformation or creep effects in the contact region. These effects would result in changes in the current. It would be a formidable task to predict the effect on current since A_s and σ enter into the current relationship in a com-plex manner, and also the effect of the non-elastic behavior on A_s and σ is not known.

58

It can be concluded, however, that if a material which is harder

than silicon is used as the pressure plate, then any plastic deformation

or creep would be limited to the silicon needle. This should result

in an increased sensitivity of the needle sensor. Although time did

not permit the testing of the hypothesis in accelerometers, it was

checked in a force transducer. The silicon pressure plate was replaced

by gold plated quartz with the result that the hysteresis effect was

eliminated within the measurement capability of the measuring instru-

ments which was on the order of 1%.

Section IV

OTHER TRANSDUCERS

4.1 Introduction

The basic silicon needle sensor can be used as the sensory element in a host of transducer applications. The configurations for employing the silicon needle sensor to measure such parameters as force, displacement and pressure are limited only by the imagination. Several of the more conventional configurations are discussed in the following paragraphs.

4.2 Direct Coupled Force and Displacement Transducer

The simplest method for utilizing the needle sensor to measure force and displacement is the direct coupled device. Here, force is applied directly to the needle. Figure 4-1 shows a sketch of such a device. The spring is used to produce a dc biasing force on the needle sensor. This device consists of a cylinder, a piston, which fits closely into the cylinder, and a rod through which the displacing force is coupled to the piston. The needle sensor is held by the piston so as to receive the applied force.

The relationship between applied force and displacement is given by

$$\Delta X = \frac{FL_r}{E_r A_r} + \frac{FL_p}{E_p A_p} + \Delta y_n \ , \tag{4.1}$$

where ΔX = Total displacement at end of connecting rod.

F = Displacing force.

E_r, E_p = Moduli of elasticity of connecting rod and piston.
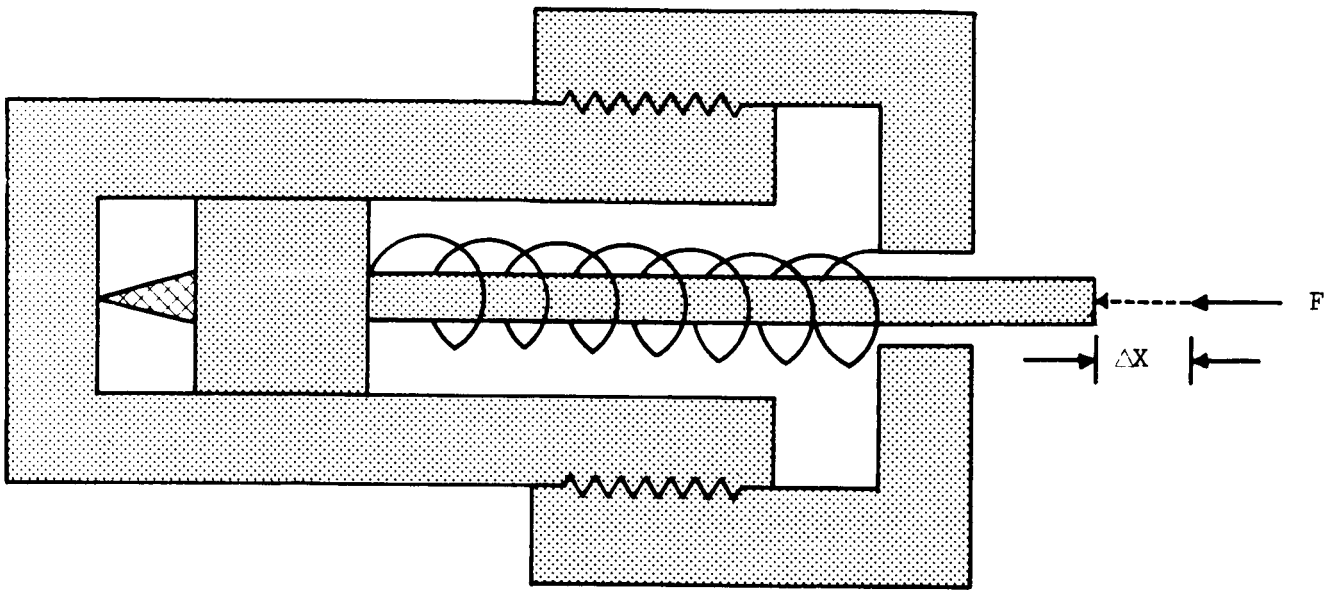
Fig. 4-1. Piston Type Displacement Transducer.

L_r, L_p = Length of rod and piston.

A_r, A_p = Cross-sectional area of rod and piston.

Δy_n = Displacement of needle.

The displacement of the needle is given in Section 2.4.

The disadvantage to this design is the friction in the cylinder.

Also, this design is susceptible to shock breakage of the needle.

4.3 Cantilever Transducer

The cantilever beam concept is readily adaptable for use with the silicon needle to measure force and displacement. The basic configuration is shown in Fig. 4-2. If F_a is the applied force on the end of the beam, then the force on the needle (F_n) is

$$F_n = F_a \frac{(3L - a)}{2a} , \qquad (4.2)$$
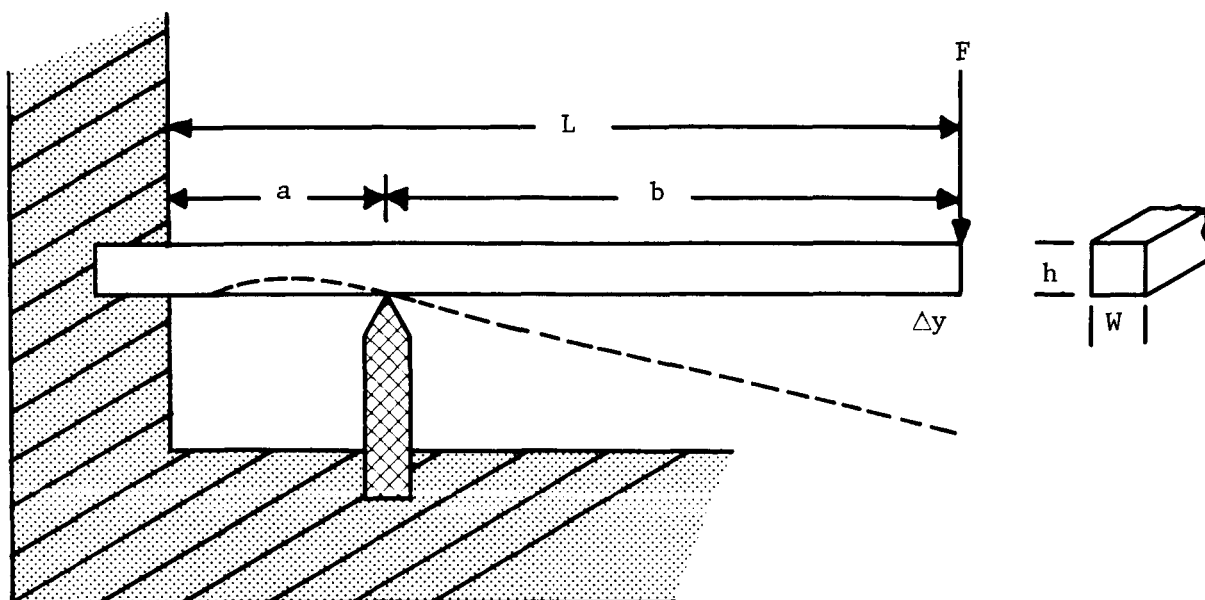
Fig. 4-2. Sketch of Cantilever Beam Force Sensor.

where L is the length of the beam and a is the distance from the canti-
lever to the needle. The deflection of the end of the beam, Δy, is

$$\Delta y = \frac{4F_a}{EWh^3} (L^3 - (3L - a)^2 a/4) , \qquad (4.3)$$

where E is Young's modulus of the beam, W is the width of the beam, and
h is the beam height. Equation (4.3) neglects the compression of the
needle. This assumption will be good for thin beams where Δy is in the
micron range.

For the case of short thick beams, the compression of the needle
must be accounted for. In this case, if Δy_n is the displacement of

the needle, then

$$F_a = \frac{2F_n a - WEh^3 \Delta y_n / 2a^2}{3L - a} , \qquad (4.4)$$

The displacement Δy_n is of course a function of F_n as discussed in Section II. The deflection Δy can be solved by combining Eq. (4.4) with the following expression

$$\Delta y = \frac{4F_a L^3}{EWh^3} - \frac{2F_n a^2}{EWh^3} (3L - a) . \qquad (4.5)$$

There are several advantages of the cantilever beam over direct coupled needle sensors. As can be seen from Eq. (4.2), there is a mechanical advantage which amplifies the force being measured. Amplifications as high as 10 can be practically realized. Also, the needle position can be interchanged with the force application position to give a mechanical disadvantage. The latter can be mathematically described by simply interchanging F_a with F_n in Eq. (4.2).

A second advantage of the cantilever beam arrangement over direct coupled transducers is its ability to withstand mechanical shock. Shock initiated at the end of the beam is attenuated to some extent before it reaches the needle sensor. Care must be exercised in the design, however, to make sure there is no side motion on the beam.

The practical design of a cantilever beam force transducer requires that the physical dimensions of the beam be considered as well as the needle sensitivity and applied force. As a force transducer, Eq. (4.2) indicates an independence from beam dimensions. There are certain practical limits to an acceptable displacement of the beam.

Also, a practical transducer requires some means for applying a dc biasing force on the needle since a needle sensor requires an approximate stress of 10^9 dynes/cm^2 before the electrical characteristics are altered. One method for applying the dc biasing stress is shown in Fig. 4-3. There are, of course, many other ways to create the dc biasing force.

A typical needle sensor has a force-sensitive range from a threshold of approximately 2 grams to a maximum of 30 grams. As a design example, assume that it is desired to design a cantilever force transducer to measure forces over the range of 0 to 5 grams. As can be seen from Eq. (4.2), a ratio of 1/3 for a/L will give 20 grams on the needle for
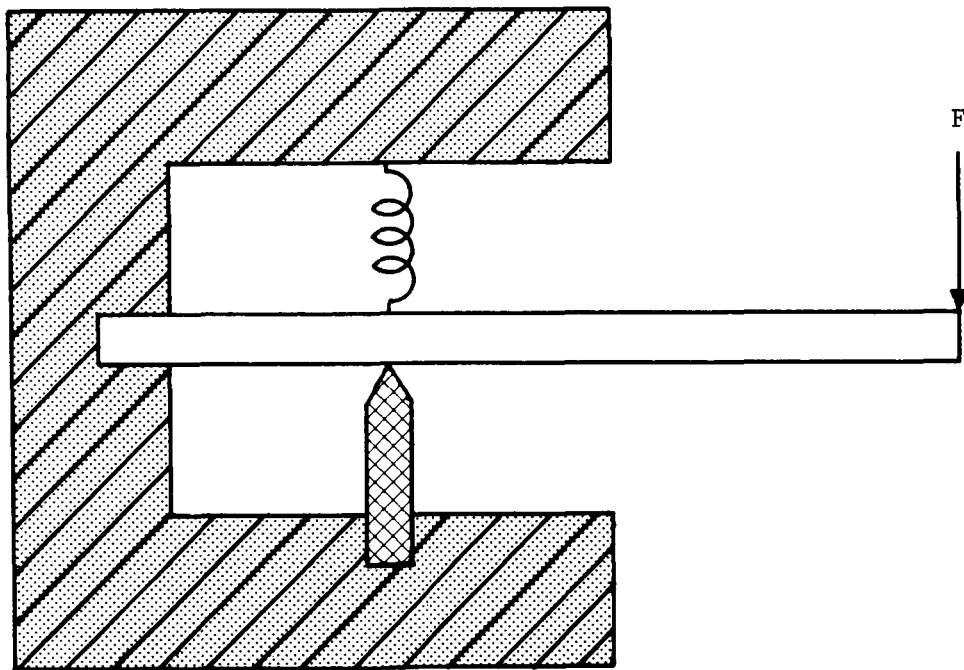
Fig. 4-3. Cantilever Transducer with a dc Biasing Force.

5 grams on the end of the beam. A dc bias load of 5 grams will put the needle in the sensitive range. This will give the needle an operating range of 5 to 25 grams.

As a design example of a cantilever displacement transducer, suppose it is desired to measure displacements over the range 0 to 40 microns. Further, assume that the maximum force that is available without interfering with the system to be measured is 1.5 grams. Consider a stainless steel beam with the following parameters:

$$L = 1/2 \text{ in.}$$
$$W = 1/8 \text{ in.}$$
$$h = .010 \text{ in.}$$
$$a = 0.05 \text{ in.}$$
$$E = 29 \times 10^6 \text{ psi .}$$

From Eq. (4.2),

$$\Delta y = (43.2 \text{ microns/gm}) \ F_a .$$

A one-gram load will give 43.2 microns in displacement which meets the displacement requirements. The one gram load will result in a force of 30 grams on the needle. The displacement of the needle is negligible in these examples.

4.4 Pressure Sensors

These are basically two methods for fabricating pressure transducers which utilize the piezojunction phenomenon. Both methods utilize a diaphragm to transmit the force to the p-n junction device. The first method utilizes the needle sensor in direct contact with the diaphragm while the second method utilizes a coupling pin between

the diaphragm and the p-n junction device. Both methods are shown schematically in Fig. 4-4.

The diaphragm in these pressure sensors can be made from such materials as stainless steel, quartz or silicon. The latter has proved to be one of the best materials to use in the pressure sensor in which the needle sensor is used.

The deflection at the center of a clamped diaphragm Δy_D under a uniform pressure is

$$\Delta y_D = - \frac{3P \ (m^2 - 1)a^4}{16Em^2t^3} \quad \text{(edges fixed) ,} \quad (4.6)$$

where P = Pressure per unit area.

  m = reciprocal of Poisson's ratio.

  a = radius of diaphragm.

  E = Young's modulus of diaphragm.

  t = diaphragm thickness.

The deflection at the center of a clamped diaphragm Δy_n under a point load is

$$\Delta y_n = \frac{3F \ (m^2 - 1)a^2}{4\pi Em^2 t^3} , \quad (4.7)$$

where F is the point load. The above equations are good for diaphragm deflections less than half the thickness of the diaphragm.

The force on the needle or coupling pin (where the radius of the coupling pin is $\ll$ a) is

$$F = \frac{P\pi a^2}{4} . \quad (4.8)$$
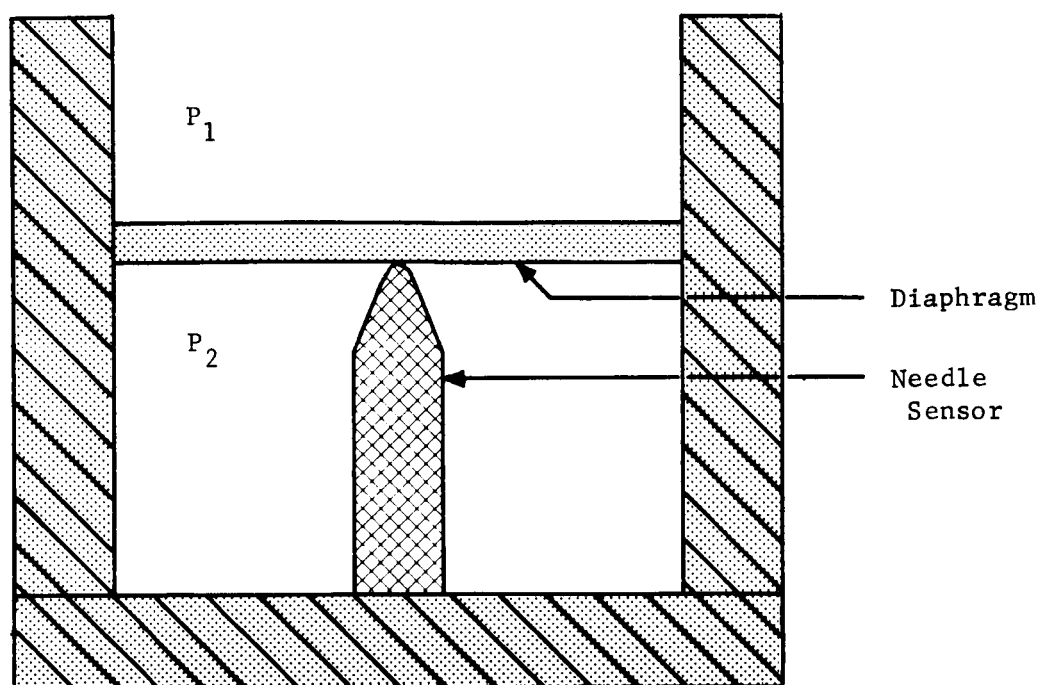
(a) Silicon Needle Pressure Sensor
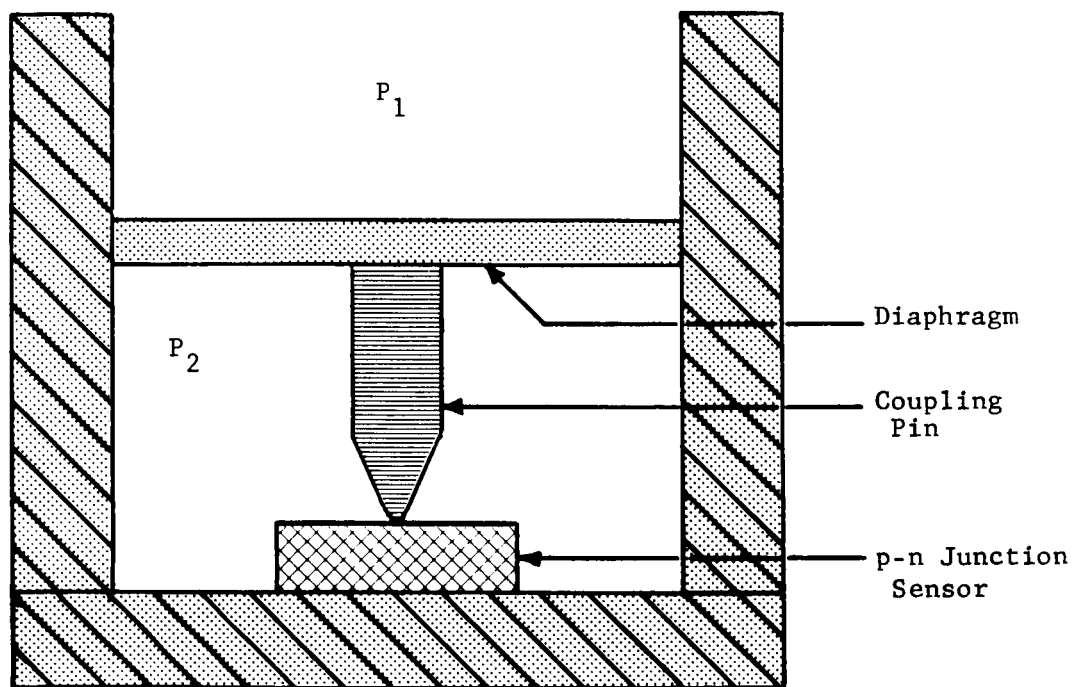
(b) Coupling Pin Pressure Sensor

Fig. 4-4. Piezojunction Pressure Sensors.

If the diaphragm is not too large in radius and is reasonably thick, it can be used to apply the dc bias force needed on the p-n junction sensors. For example, if a stainless steel diaphragm is used with t = 2 mils, a = 0.2 in., a biasing force of 10 grams can be placed on the needle by forcing it against the diaphragm with only an 0.8 mil deflection. In this example, a pressure differential across the diaphragm of 1 psi will apply a force of 14.3 grams on the sensor which is in the proper range for the needle sensor to operate.

Pressure transducers utilizing both the coupling pin and the needle sensor have been fabricated and tested. One of the major advantages of the coupling pin method is that complicated multijunction sensors can be used whereas in needle sensors only one junction is presently available.

A differential pressure transducer of the coupling pin type has been fabricated in which the p-n junction sensor was a four-layer switch. The switch was made using the planar process with a single window as described in Section II.

The transducer consists of a brass housing, a brass diaphragm, inlet and outlet ports, a stainless steel needle, and a planar four-layer switch in chip form. These components were used to form three major assemblies which were then integrated into the final transducer shown schematically in Fig. 4-5.

The diaphragm was cut from 2 mil brass shim stock and fitted to one end of the brass inner housing. The other end of this inner housing was drilled to accommodate an insulated post containing the sensing element and to accommodate the low pressure port. The outer
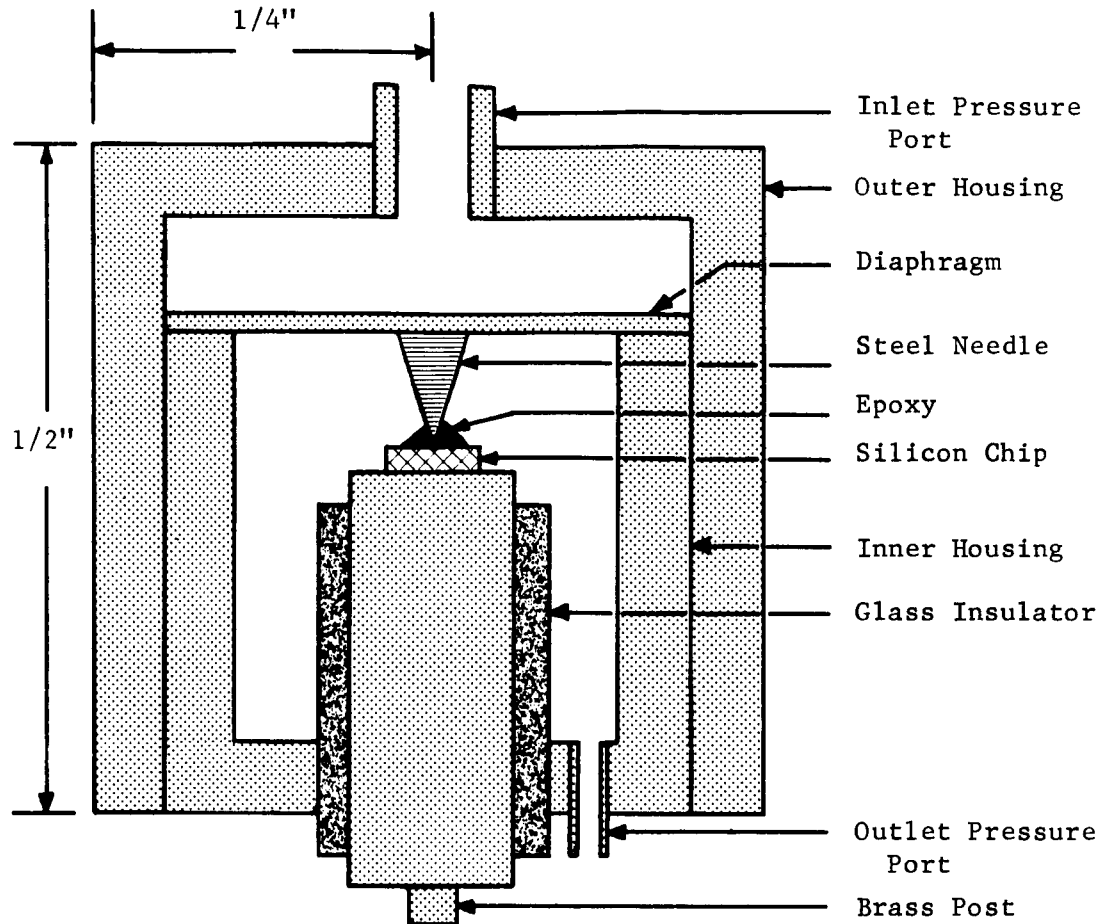
Fig. 4-5. Sketch of Four-Layer Switch Pressure Transducer.

housing was machined so that the inner housing (containing the diaphragm) would fit closely into it. One end of the outer housing was fitted with a brass inlet tube. The brass diaphragm and the inlet and outlet ports were all connected using lead soldering techniques.

The sensing element was made by mounting the chip containing the four-layer switch onto one end of a brass post which was machined to fit closely into a 3/8 inch O.D. glass tube. The chip was held in place by conductive epoxy. The end of the brass post opposite the silicon chip was machined into a terminal for electrical connection.

The purpose of the glass was to electrically insulate the post from the remainder of the transducer.

A steel needle was epoxied to the silicon chip to transmit the force from the diaphragm to the four-layer switch. Armstrong resin epoxy was used.

The first step in the integration of these assemblies was to place the inner housing into the outer housing leaving a small chamber between the inlet port and the diaphragm. The two housings were bonded together by resin epoxy. The sensing element was then placed into the glass insulator and epoxied in place. Conductive epoxy was placed on the bare end of the steel needle and the complete sensing element assembly lowered into the inner cylinder until electrical contact was made between the needle and the diaphragm. Once contact was made, an initial force was produced by lowering the sensing element assembly still further into the inner housing. A photograph of the device is shown in Fig. 4-6.
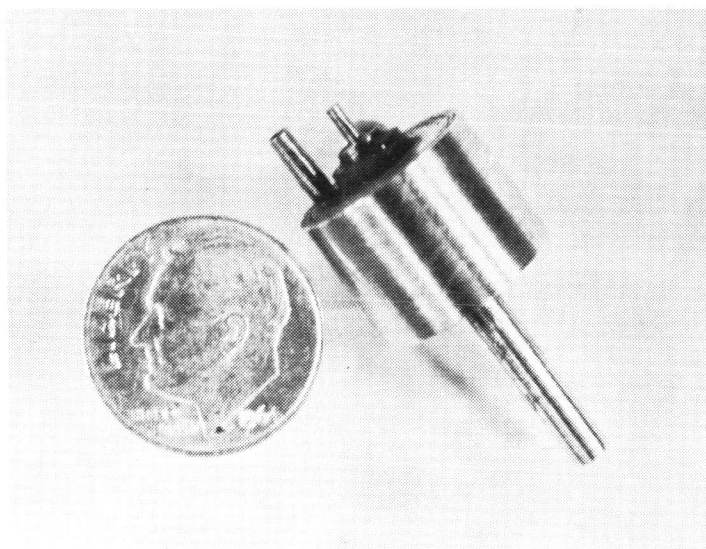
Fig. 4-6. Four-Layer Switch Pressure Transducer.

The completed pressure transducer was calibrated by connecting the four-layer sensing element in a relaxation oscillator circuit (see Section V) and applying a known pressure to the inlet port. The outlet, or low pressure port, was vented to the atmosphere. The center or zero pressure frequency of the oscillator was 91 kc. A frequency change of 20 kc was obtained with a pressure change of 0 to 4 psi. The calibration curve obtained in this manner is shown in Fig. 4-7.
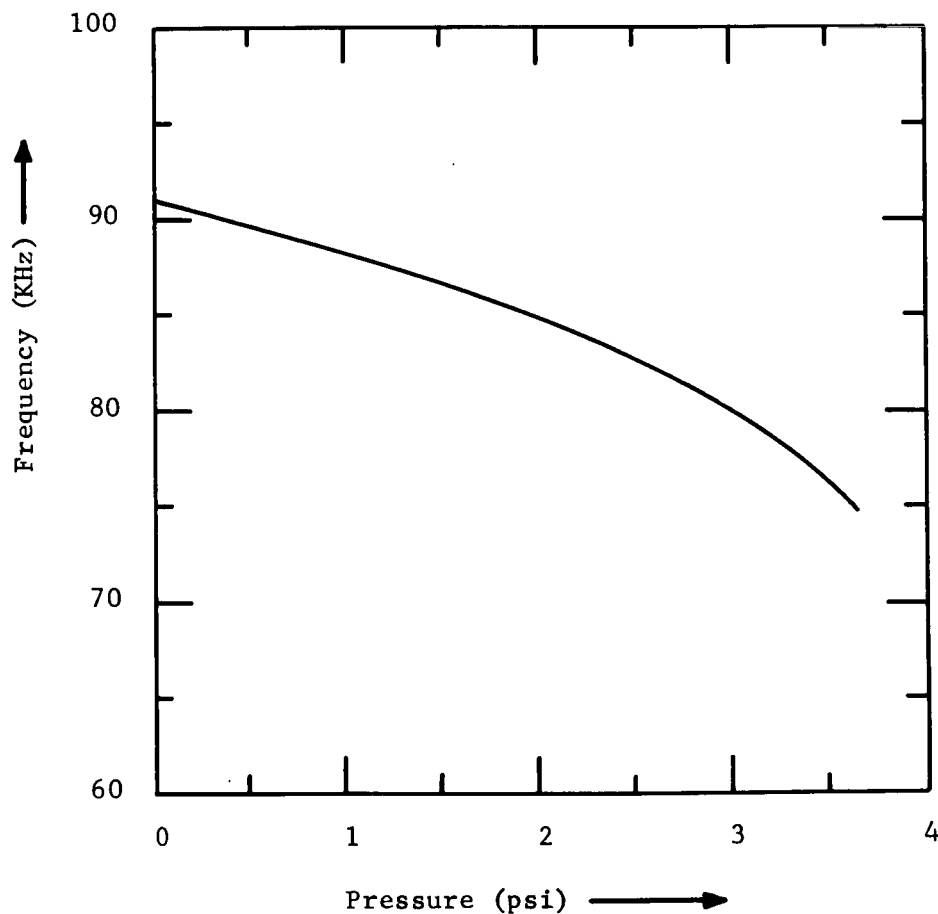
Fig. 4-7. Calibration Curve of Four-Layer Switch Pressure Transducer.

Silicon needle pressure transducers have also been fabricated and tested. These transducers utilized a silicon diaphragm with a thickness of 2.5 mils. The diaphragm was clamped between a rubber O ring and a brass housing. The silicon needle sensor was mounted in a mechanical screw and set in place against one side of the diaphragm. Electrical contact was made by evaporating gold on the silicon diaphragm to which one side of the needle junction made contact. The needle shank was the other contact. The dc biasing force was applied by putting a vacuum on the side of the diaphragm opposite the needle and then setting the needle against the diaphragm by means of the screw. When the vacuum was released, the diaphragm applied the dc bias force. The diameter of the diaphragm was 12 mm. Figure 4-8 is a plot of the sensor current as a function of pressure. The voltage was held constant across the sensor.
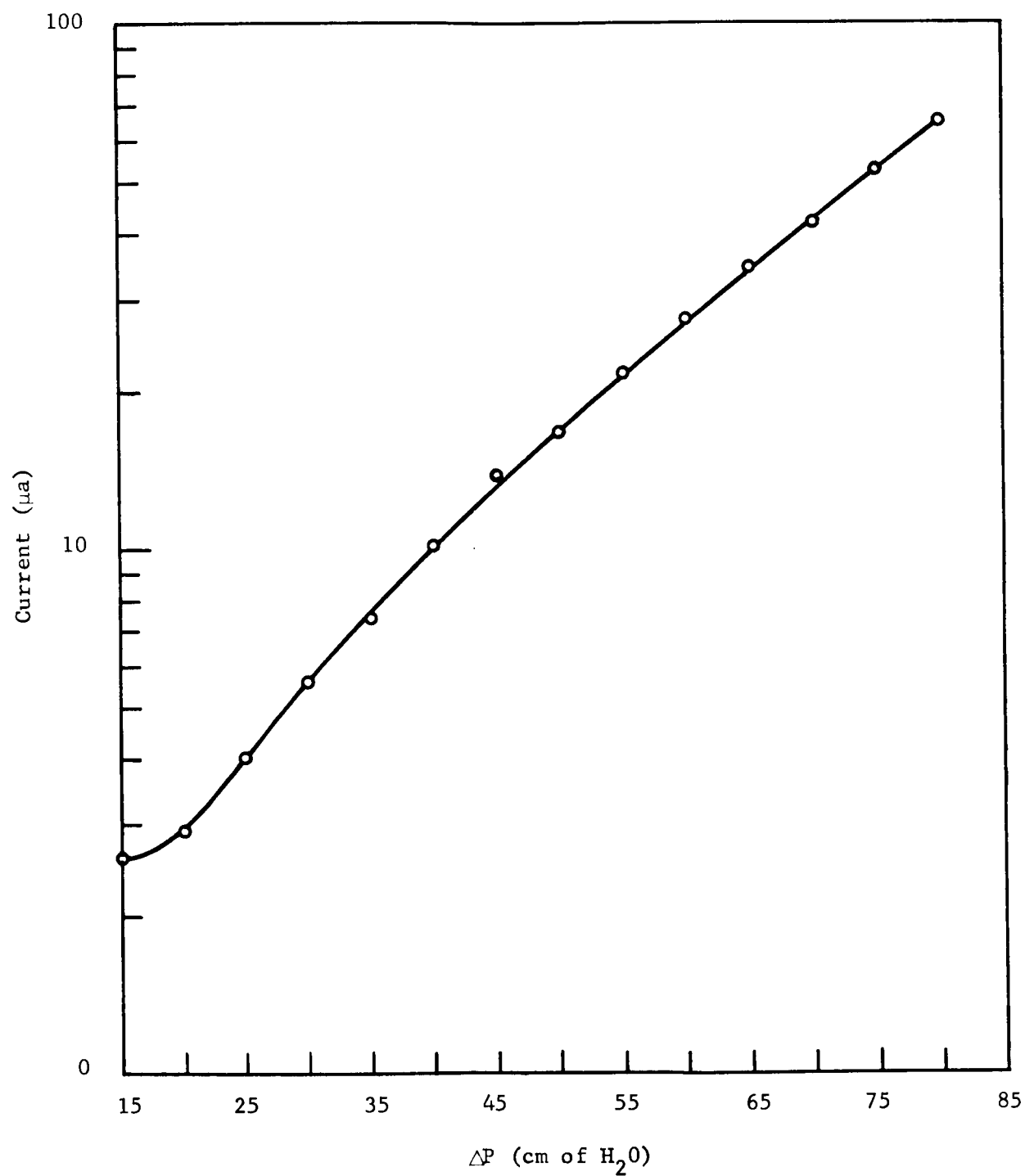
Fig. 4-8. Current in the Reverse Biased Mode as a Function of Pressure for a Silicon Needle Pressure Transducer.

Section V

Read-Out Circuitry for Piezojunction Sensors

5.1 Introduction

Circuitry used in the past for acquiring data from laboratory accelerometers and piezojunction devices was devised specifically for the purpose of evaluating these devices under laboratory conditions. The measuring techniques were conventional and implemented conventional laboratory equipment. The development of accelerometers and other transducers has produced a need for considering circuits which can be used to produce a read-out for the devices in other than laboratory applications. Piezojunction sensors are solid state devices and hence are compatible with microelectronic techniques and devices.

5.2 Differential Amplifier

A simple differential amplifier was devised which produced a 0 - 5V dc output in proportion to a stress input of 5×10^9 to 10^{10} dynes/cm^2 into a p-n junction sensor. Two similar reverse biased diodes were used - one in each leg of the differential amplifier. One diode was the sensing element of an accelerometer and the other was a reference element. The diodes formed a part of a biasing network for the amplifier stages. The zero point is balanced by a variable resistor in the reference stage. A schematic diagram of the amplifier is shown in Fig. 5-1.

5.3 Four-Layer Switch Oscillator

Figure 5-2 is a schematic diagram of a relaxation oscillator in which the frequency is dependent on the input acceleration. The

(73)
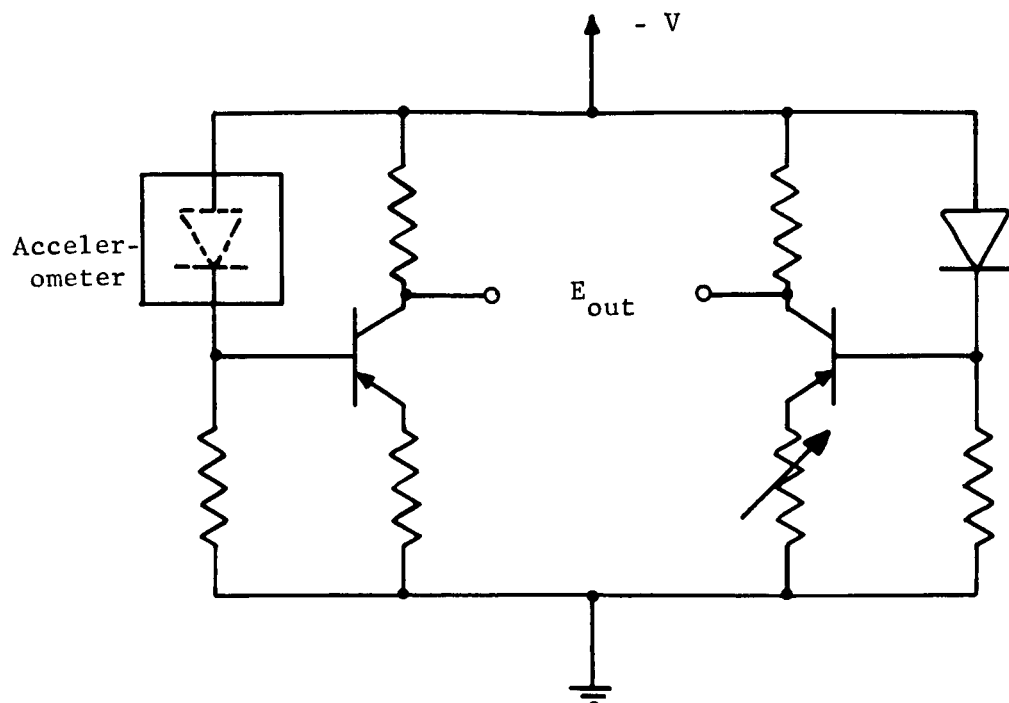
Fig. 5-1. Accelerometer Differential Amplifier.
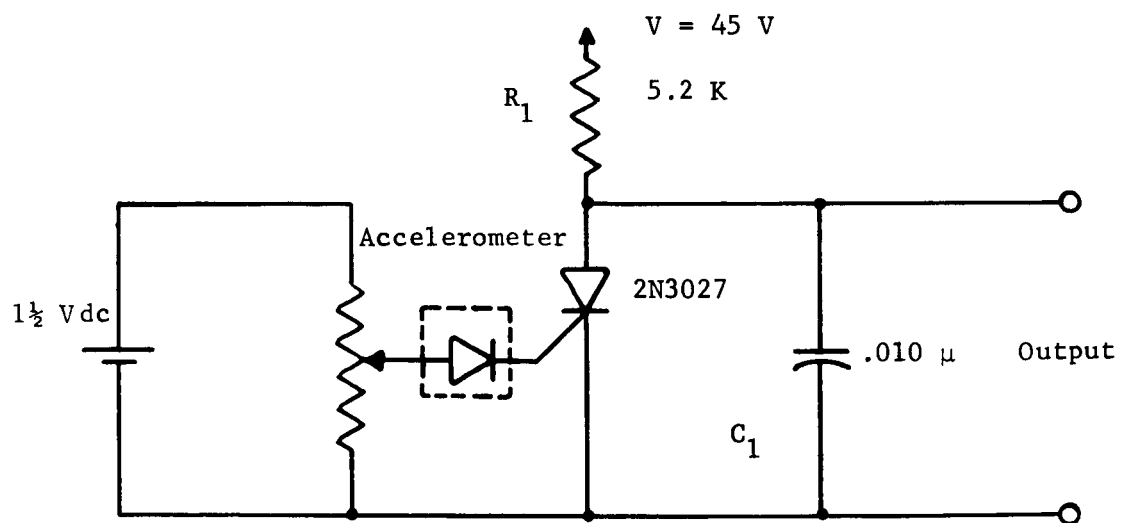
Fig. 5-2. Four-Layer Switch Oscillator Circuit.

active element in the oscillator is a 2N3027 four-layer switching

device. The frequency of oscillation depends on the gate current to

the device. An accelerometer is placed in series (such that the

acceleration sensing junction device is forward biased) with the gate

supply. An increase in acceleration allows a rise in gate current and

results in an increase in frequency of oscillation. The oscillator

frequency is controlled by the R_1C_1 time constant and R of the sensor.

Figure 5-3 shows two photographs of the output waveform of the circuit,

(a) with no acceleration and (b) under acceleration. A change in both

the amplitude and frequency of the waveform can be easily detected

in these photographs. However, the change in frequency is the most

advantageous measure of input acceleration due to the fact that it can

be more easily transmitted without distortion.

The most critical disadvantage of the above circuit has been

its instability.

5.4 Unijunction Oscillator

A similar circuit, shown in Fig. 5-4, was designed which utilizes

a unijunction transistor as the active element in a relaxation

oscillator. The accelerometer acts as an acceleration sensitive

resistor in the RC timing circuit of the oscillator. It is this series

RC circuit which determines the oscillator frequency. This oscillator

is more stable with frequency than the previous circuit. This is

partially due to the fact that the frequency of oscillation does not

depend directly on a switching level of the unijunction device but

is determined by an external circuit. Figure 5-4 is a schematic of the

unijunction transistor circuit.
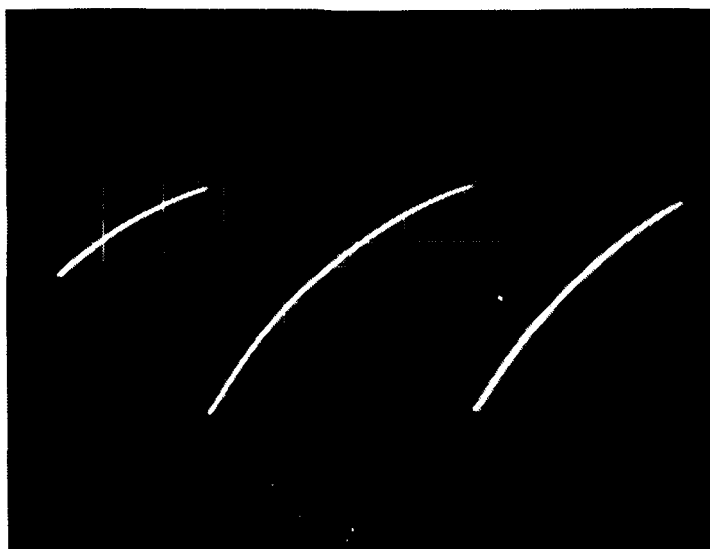
(a)   No Acceleration
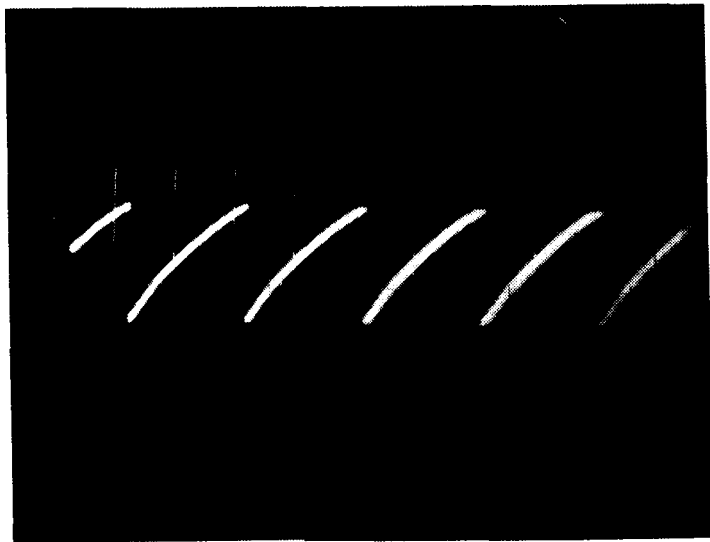
(b)   Under Acceleration

Fig. 5-3.   Photographs of Four-Layer Switch Oscillator Waveforms Under
            Normal and Higher-Than-Normal Accelerations.
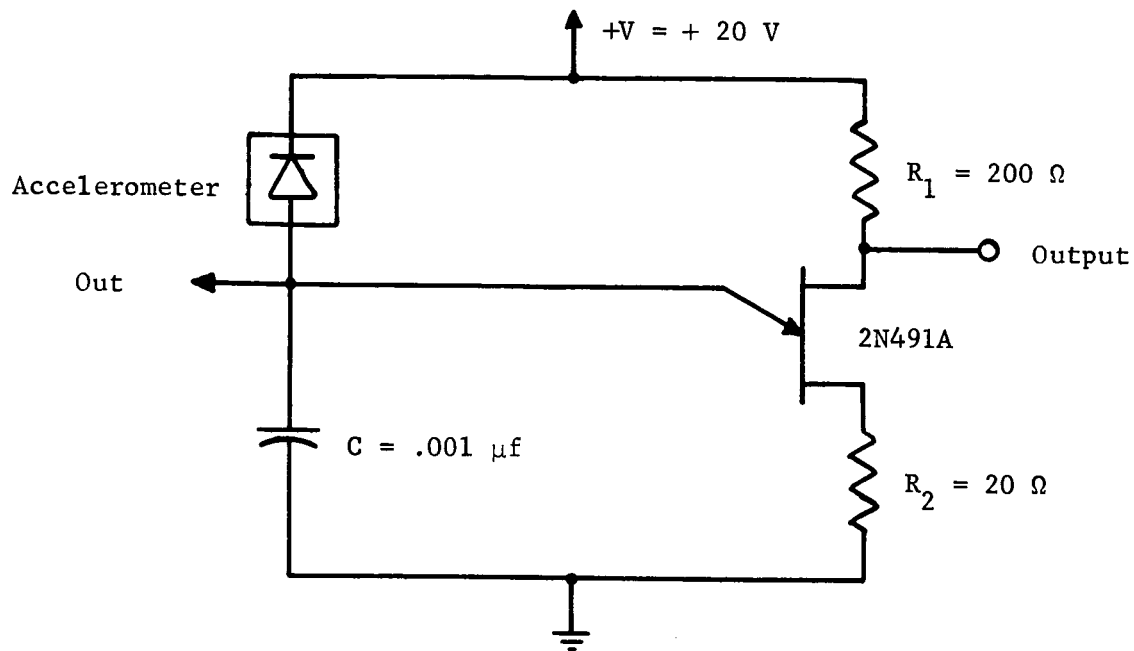
Fig. 5-4. Unijunction Transistor Oscillator Circuit.

5.5 Summary

The circuits described above were developed simply to determine their feasibility and are by no means optimum. There has been no attempt made to environment harden any of the circuits, although the differential amplifier is somewhat inherently temperature compensated. The simplicity of the circuits reflect the lack of difficulty associated with utilizing the piezojunction accelerometers in a practical application. Probably most significant is the ease with which a frequency modulated read-out is obtained.

study have shown that, in fact, the piezojunction phenomenon can be used as the sensory phenomenon in accelerometers, force, displacement, and pressure transducers. No finalized designs were made, the object here being a study of feasibility of such devices.

Considerable effort has been placed on the demonstration of the usefulness of the piezojunction effect in accelerometers. The double-diaphragm type device was found to be the best configuration studied for use with the needle sensor. These devices showed resonant frequencies greater than 3 KC and are capable of measuring both ac and dc accelerations. Laboratory devices were fabricated with responses covering the range from $\pm$ 1 g to $\pm$ 100 g.

Several techniques were developed for providing a digital output for the piezojunction sensors. The most prominent were the four-layer oscillator and the unijunction oscillator.

There still remains some questions that need to be answered in the application of the piezojunction effect in transducers. One of these is the hysteresis effect as exhibited by the laboratory accelerometers. Although quartz pressure plates appeared to eliminate the effect, it remains to be proven in the accelerometers themselves. Another problem that needs further study is the fragile nature of piezojunction transducers. Methods of preventing overstress and breakage of the sensors need to be investigated. Temperature and long term stability studies of transducers also need to be performed if the devices are to be applied with success and confidence.

The limitations that have been encountered in applying the piezo-junction effect in transducers do not appear to be beyond elimination

Section VI

CONCLUSIONS AND RECOMMENDATIONS

This study has shown that mechanical stress applied to the junction area of p-n junction devices can cause large changes in the electrical characteristics of the devices. The underlying mechanism has been found to be a decrease in the effective energy gap of the semiconductor crystal. Analytical expressions have been developed for the effects of stress on the electrical characteristics of diodes, transistors, and four-layer switches. The piezojunction effect showed up in many ways depending on the device configuration and mode of operation. Of the many devices that were studied, the single-junction diode with stress applied to the entire junction area has emerged as one of the best configurations for use in transducer applications.

The silicon needle sensor was conceived and the techniques for fabrication were developed. The needle sensor has the big advantage of not requiring a delicate alignment operation between the stressing arrangement and the junction to be stressed. It also offers the possibility of drastic size reductions in sensors based on the piezo-junction effect. One of the problems associated with the application of the needle sensor has been the reduction of the size of the mechanical configurations so that they were not orders of magnitude greater than the sensor size.

A variety of transducers utilizing both the needle sensor and the indentor point method have been investigated. Many of these transducers were fabricated in the laboratory and evaluated. The results of the

(79)

in practical devices. The phenomenon offers the possibility of a whole new class of transducers that are limited only by one's imagination. The major attributes of transducers based on the piezojunction effect are small size, low power, high sensitivity, high resonant frequency, sensitive to both ac and dc stimuli, and compatibility with micro-electronic circuits and techniques.

LIST OF REFERENCES

1   R. J. Roark, <u>Formulas For Stress And Strain</u>, 9th Ed., McGraw-Hill Book Company, New York, New York (1965).

2   K. Bulthuis, Phillips Res. Repts., <u>21</u>, (1966) pp. 85 - 103.

3   W. Rindner and I. Braun, J. Appl. Phys., <u>34</u>, (1963) pp. 1958 - 1970.

4   S. Timoshenko and J. N. Goodier, <u>Theory Of Elasticity</u>, McGraw-Hill Book Company, New York (1951).

5   H. Hertz, Miscellaneous Papers. Also, Love, <u>Theory of Elasticity</u>, 3rd Ed., Cambridge University Press, (1928).

Appendix A

STRESS EFFECTS ON p-n-p-n DIODES

Appendix A

STRESS EFFECTS ON p-n-p-n DIODES

The basic p-n-p-n structure with two gate connections is shown in Fig. A-1. In the low current "off" state of the device, the center junction is reverse biased while the other two junctions are slightly forward biased. The current through the center junction is composed of three components.

(1) The normal reverse bias current of a p-n junction, $I_{CO}(V)$.

(2) The fraction of the hole current at junction 1 which is collected at the center junction , $M_n \alpha_N I_C$.

(3) The fraction of the electron current at junction 3 which is collected at the center junction, $M_p \alpha_P I_E$.

The quantities α_N and α_P are the dc alphas of the n-p-n and the p-n-p sections treated as transistors. M_n and M_p are the multiplication
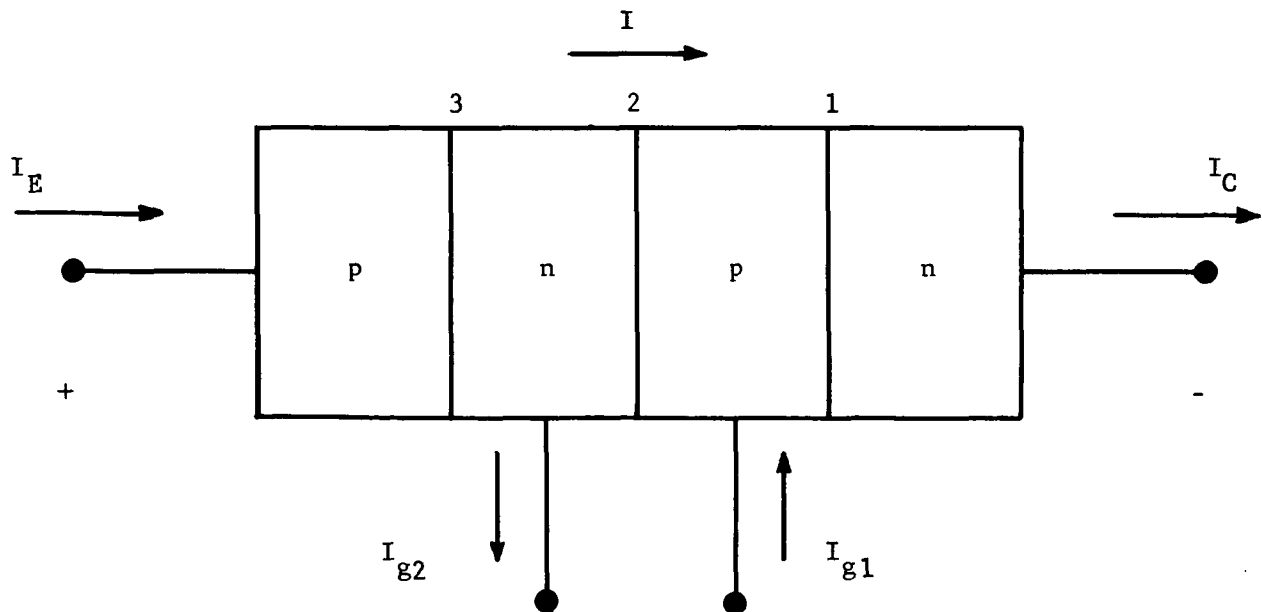
Fig. A-1. Basic p-n-p-n Structure.

factors for electrons and holes. The current at the center junction
is then

$$I = I_{CO}(V) + M_n(V)\alpha_N I_C + M_p(V)\alpha_P I_E \ . \qquad (A.1)$$

The currents I_{g1} and I_{g2} are normally taken as the independent vari-
ables and the current equation written as

$$I = I_{CO}(V) + M_n(V)\alpha_N(I + I_{g1}) + M_p(V)\alpha_P(I + I_{g2}) \qquad (A.2)$$

Equation (A.2) can be solved for I and gives

$$I = \frac{I_{CO}(V) + M_n(V)\alpha_N I_{g1} + M_p(V)\alpha_P I_{g2}}{1 - M_n(V)\alpha_N - M_p(V)\alpha_P} \qquad (A.3)$$

The alphas, α_N and α_p, are functions of the currents at junctions 1
and 3, respectively; i.e.,

$$\alpha_N = \alpha_N(I + I_{g1}) \ ,$$
$$\qquad (A.4)$$
$$\alpha_P = \alpha_P(I + I_{g2}) \ .$$

It is obvious from Eq. (A.3) that I becomes very large as

$$M_n(V)\alpha_N(I + I_{g1}) + M_p(V)\alpha_P(I + I_{g2}) \to 1 \ . \qquad (A.5)$$

The I-V characteristic described by Eq. (A.3) is a double-valued func-
tion of I because of the dependence of α_N and α_p on I. A typical
characteristic is shown in Fig. A-2. The coordinates of the switching
point separating the high and low impedance states are I_S, V_S. A
negative resistance region occurs between I_S, V_S and I_h, V_h. The
switching point can be shown to occur when

$$M_n(V_S)\alpha_{no}(I_S + I_{g1}) + M_p(V_S)\alpha_{po}(I_S + I_{g2}) = 1 \ , \qquad (A.6)$$
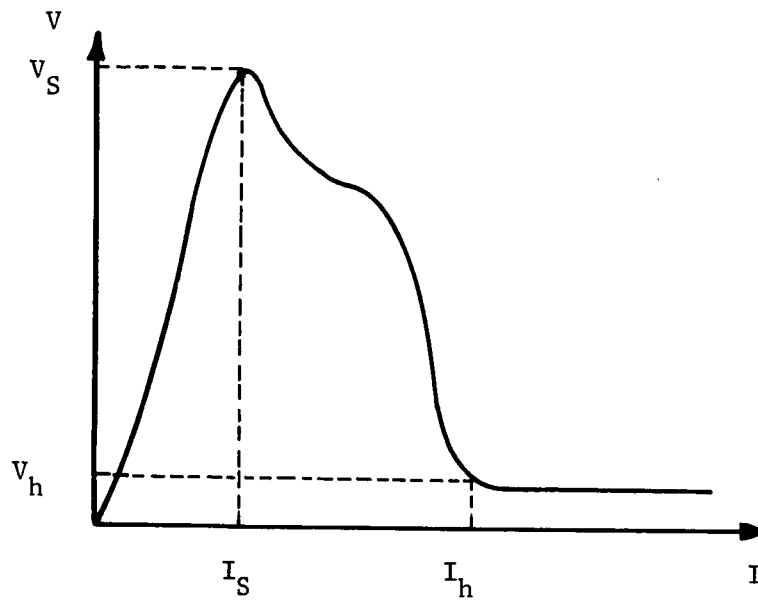
Fig. A-2. Typical I-V Characteristic of p-n-p-n Diode.

where α_{no} and α_{po} are the small signal alphas of the n-p-n and p-n-p regions. Equation (A.6) combined with the equation

$$I_S = \frac{I_{CO}(V_S) + I_{g1}M_n(V_S)\alpha_N(I_S + I_{g1}) + I_{g2}M_p(V_S)\alpha_p(I_S + I_{g2})}{1 - M_n(V_S)\alpha_N(I_S + I_{g1}) - M_p(V_S)\alpha_p(I_S + I_{g2})}, \quad (A.7)$$

determines the switching coordinates I_S and V_S. For an alpha which increases with current, it can be shown that

$$\alpha_{no} \geq \alpha_N \ . \tag{A.8}$$

The case of $I_{g1} = I_{g2} = 0$ is of interest here, and gives

$$I = \frac{I_{CO}(V)}{1 - M_n(V)\alpha_N(I) - M_p(V)\alpha_p(I)} \ . \tag{A.9}$$

The function $I_{CO}(V)$ has the approximate form:

$$I_{CO}(V) = I_o \left(\frac{V}{V_B}\right)^{1/n} M(V) \ , \qquad (A.10)$$

where $M(V)$ is the multiplication factor for the center junction, V_B is the breakdown voltage, and I_o is the reverse bias current at the center junction with no multiplication extrapolated to the breakdown voltage. This is illustrated in Fig. A-3. Normalizing the voltage in this manner at V_B is done for convenience only. The parameter n is typically between 2 and 3. The multiplication factor is assumed to be of the form

$$M(V) = \frac{1}{1 - \left(\frac{V}{V_B}\right)^m} \ . \qquad (A.11)$$

Fig. A-3. Reverse Bias Current vs Reverse Voltage.

For most applications of p-n-p-n diodes, high switching voltages are desirable and consequently most commercial devices have $\alpha_{no}(I_o) + \alpha_{po}(I_o) \ll 1$. For these devices, the threshold voltage is essentially the breakdown voltage. Junction breakdown voltage is fairly insensitive to stress and thus type A devices are not good diodes for stress transducer applications. On the other hand, the switching voltage of type B devices is very sensitive to changes in the alpha of the various sections and consequently devices of this type are suited for stress transducer elements. It should be noted that a type A device can be converted to a type B device by applying a gate current to a p-n-p-n switch and thereby increasing the alpha of one of the transistor sections. Further discussion is limited to devices for type B where the threshold voltage is less than the breakdown voltage. For this case, multiplication is relatively unimportant and

$$I \simeq \frac{I_{CO}(V)}{1 - \alpha_N(I) - \alpha_P(I)} \; . \tag{A.14}$$

The current at which switching occurs is determined by the equation:

$$\alpha_{no}(I_S) + \alpha_{po}(I_S) \simeq 1 \; . \tag{A.15}$$

A model for the p-n-p-n diode with a small area under stress is shown in Fig. A-5. The stress is assumed to be applied uniformly over an area A_S. In an analysis similar to that leading to Eq. (A.1), the total current is

$$I = I_1 + I_2 = I_{CO}(V) + \alpha_P I_{a1} + \alpha_N I_{C1}$$

$$+ \alpha_P' I_{a2} + \alpha_N' I_{C2} \; . \tag{A.16}$$
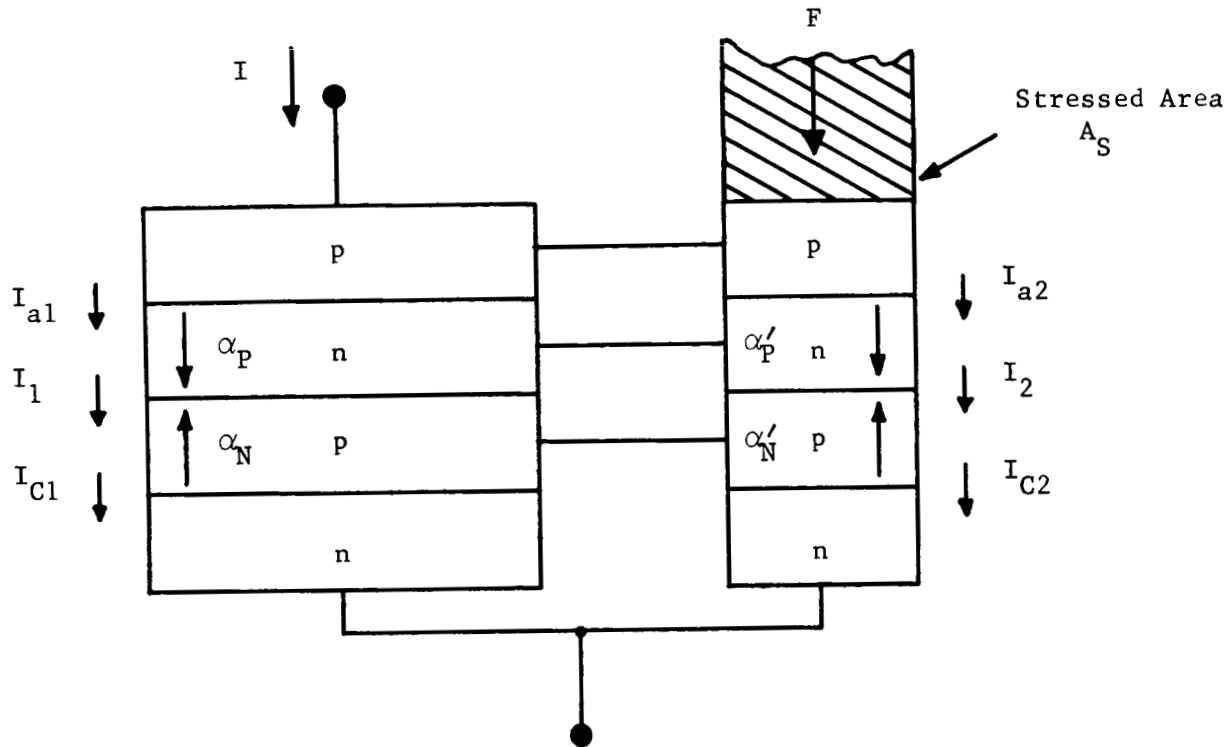
Fig. A-5. Model for Stressed p-n-p-n Diode.

This can be solved for the total current as

$$I = \frac{I_{CO}(V)}{1 - \alpha_P \dfrac{I_{a1}}{I} - \alpha_N \dfrac{I_{C1}}{I} - \alpha_P' \dfrac{I_{a2}}{I} - \alpha_N' \dfrac{I_{C2}}{I}} \; . \qquad (A.17)$$

For the most general stress case, the alphas depend upon the stress and current densities in the various regions. Also, the ratios I_{a1}/I, I_{C1}/I, etc., can depend upon the stress as well as total current.

The effect of stress on the alpha of a p-n-p transistor can be calculated by considering the current components:

$$I_{pn} = I_1 \gamma(\sigma_n) \exp(qV/kT) \; ,$$

$$I_{np} = I_2 \gamma(\sigma_p) \exp(qV/kT) \; , \qquad (A.18)$$

$$I_{sc} = I_3 \sqrt{\gamma(\sigma_J)} \exp(qV/2kT) \; ,$$

where $\gamma(\sigma_n)$, $\gamma(\sigma_p)$, $\sqrt{\gamma(\sigma_J)}$ give the effects of stress on the current components. The current I_{SC} is the space charge region generation-recombination current. The total current is then

$$I = [I_1\gamma(\sigma_n) + I_2\gamma(\sigma_p)] Z^2 + I_3\sqrt{\gamma(\sigma_J)}\; Z \; , \qquad (A.19)$$

where

$$Z = \exp(qV/2kT) \; . \qquad (A.20)$$

The solution of this equation for Z is

$$Z = \frac{I_3\sqrt{\gamma(\sigma_J)}}{2[I_1\gamma(\sigma_n) + I_2\gamma(\sigma_p)]} \{[1 + 8\; I/I^*]^{1/2} - 1\} \; , \qquad (A.21)$$

where

$$I^* = \frac{2\; I_3^{\;2}\gamma(\sigma_J)}{[I_1\gamma(\sigma_n) + I_2\gamma(\sigma_p)]} \qquad (A.22)$$

The dc and small signal alphas are

$$\alpha = \frac{\beta^* I_{pn}}{I} \; ,$$

$$\alpha_o = \beta^*\; \frac{\partial I_{pn}}{\partial I} \; , \qquad (A.23)$$

where β^* is the base transport factor which is assumed to be independent of current. Evaluation of the expressions for alpha in terms of I gives

$$\alpha = \alpha_\infty \frac{1}{1 + 2\;[(1 + 8\; I/I^*)^{1/2} - 1]^{-1}} \qquad (A.24)$$

$$\alpha_o = \alpha_\infty \left[1 - \frac{1}{(1 + 8\; I/I^*)^{1/2}}\right]$$

where

$$\alpha_\infty = \beta^* \frac{I_1 \gamma(\sigma_n)}{I_1 \gamma(\sigma_n) + I_2 \gamma(\sigma_p)} \; . \qquad (A.25)$$

For the case of uniform stress across the emitter junction, i.e.,

$\sigma_n = \sigma_p = \sigma_J$, it is seen that I* and α_∞ are independent of stress. At

a given current level, then alpha is independent of stress. For non-

uniform stress across the junction, such as $\sigma_p > \sigma_J > \sigma_n$, α_∞ decreases

with stress, since $I_1 \gg I_2$. I* first increases and then decreases as

stress is increased for the above case. The behavior of alpha at a

constant current level is then a complex function of stress. At suf-

ficiently large stress alpha decreases. In the limit of $\sigma_J = \sigma_n = 0$

and $\sigma_p \neq 0$, both I* and α_∞ decrease with increasing stress.

The behavior of p-n-p-n diodes under general stress is very com-

plicated, therefore, some simple cases are considered. The first

case is that of uniform stress over the entire diode, i.e., $(A_S = A_T)$.

As discussed above, the alphas for this case depend only on the cur-

rent and not on the stress. The only stress dependent term in

Eq. (A.17) is then I_{CO}, the reverse bias current of the center junc-

tion which at the switching point gives

$$I_S = \frac{I_0 (\frac{V_S}{V_B})^{1/n} [\gamma(\sigma)]^{1/2}}{1 - \alpha_N(I_S) - \alpha_P(I_S)} \; , \qquad (A.26)$$

where the stress dependence of I_{CO} has been taken as $\sqrt{\gamma}$. Since the

alphas are independent of stress, the switching current is independent

of stress and

$$V_S = V_B \{\frac{I_S}{I_o}[1 - \alpha_N(I_S) - \alpha_P(I_S)]\}^n \ [\gamma(\sigma)]^{-n/2}$$

or   (A.27)

$$V_S \propto [\gamma(\sigma)]^{-n/2} \ .$$

Since γ increases with stress, the switching voltage is seen to de-
crease with stress.

When a small region is stressed such as shown in Fig. A-5, the
current is no longer uniform over the diode. The lowered band-gap
in the stressed area causes a larger fraction of the total current to
flow in the stressed area, and a corresponding smaller fraction of the
to al current in the unstressed area. The increased current in the
stressed area increases the alphas of the stressed area leading to
further increases in current. The increased current density in the
stressed area might be expected to give a lower switching voltage,
however, this is not necessarily the case since the total current is
that through both the stressed and unstressed areas.

Further complications arise when a small area is stressed due to
non-uniform stressing of the junctions. Stress over small areas is
normally applied by means of an indenter point which gives an atten-
uation of the stress with depth from the surface. To illustrate
this case, assume that the force is concentrated mainly on the p side
of the upper junction in Fig. A-5. For this case, stress causes an
increase in the injection of electrons into the upper p region. This
decreases α_P' and the effect can be represented as a stress dependent

gate current bypassing the upper p-n junction. This is illus-
trated in Fig. A-6. This increased current acts to suppress the
switching action just as a gate current in the same direction would
do. Thus, under these conditions, the switching voltage would increase.

As the above limiting cases illustrate, there are several com-
peting processes in operation for most practical stress conditions.
Some of these factors tend to give a decreasing switching voltage
with increasing stress while other factors have just the opposite
effect. The switching voltage can thus be expected to decrease or
increase depending on the relative importance of these factors.

$$\frac{I_g}{I} \propto (\gamma - 1) \frac{A_S}{A_T}$$
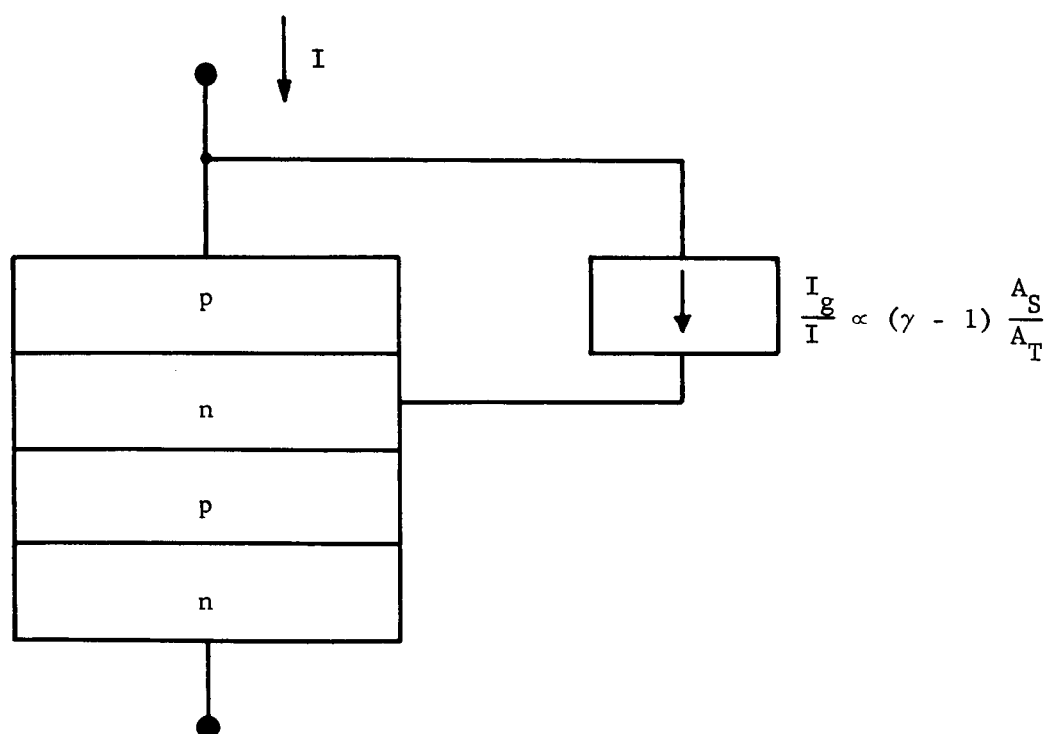
Fig. A-6. Effect of Small Area Stress with Rapid Stress Attenuation
from Stressed Surface.

The important factors are:

    (1) Non-uniform stress applied to the various junctions.

    (2) Different current densities in the stressed and unstressed areas.

While the case of a general stress has not been analyzed in detail here, the following general conclusions can be drawn. For the case where the stressed area and total area do not differ greatly, the switching voltage should decrease with increasing stress. For the case of stressing a very small area with an indenter point, the switching voltage should increase with increasing stress. The changes with stress in either case can be rather large. Practical stressing arrangements are often between these extremes and can give either increasing or decreasing switching voltage as stress is increased.

Appendix B

SOME EFFECTS OF MECHANICAL STRESS ON THE BREAKDOWN
VOLTAGE OF p-n JUNCTIONS

Appendix B

Some Effects of Mechanical Stress on the Breakdown

Voltage of p-n Junctions[*]

J. R. Hauser and J. J. Wortman
Research Triangle Institute, Research Triangle Park, North Carolina

Various factors determining stress induced changes in the break-
down voltage of Ge and Si p-n junctions are discussed. A model is
developed which accounts for the multivalley band structure of semi-
conductors such as Ge and Si. Analytic expressions are developed
for the change in breakdown voltage, $\Delta V/V_B$, as a function of a general
stress. For Si, a linear decrease in $\Delta V/V_B$ with increasing stress is
predicted. The proportionality factor is of the same order of magni-
tude as the band gap dependence on hydrostatic pressure. For Ge, an
initial increase in $\Delta V/V_B$ followed by a decrease at high stress levels
is indicated. The model is shown to be consistent with reported experi-
mental data.

I. INTRODUCTION

Several investigations have shown that mechanical stress can induce reversible changes in the electrical characteristics of p-n junctions.[1-5] In these investigations, large anisotropic stresses have been introduced into the junctions of diodes and transistors resulting in large changes in saturation currents of diodes and in current-gain for transistors. These changes have been attributed to stress induced changes in the energy band structure of the material and in particular to changes in the band gap.[6] Recent experimental investigations have shown that the reverse breakdown voltage in Si and Ge diodes is also stress sensitive.[7,8] The breakdown voltage in Si is found to be linearly related to uniaxial compression stress, while in Ge, the breakdown voltage is found to be a more complex function of the stress. The stress coefficient in Si is approximately equal to the coefficient found for band gap lowering under hydrostatic pressure. In Ge, the breakdown voltage is found to increase and then to decrease as stress increases.

The purpose here is to discuss the effects of multiple energy bands on the stress dependence of breakdown and to evaluate the order of magnitude of these stress induced changes in the breakdown voltage of Ge and Si diodes. A theoretical discussion is given which is based on Shockley's "Simple Model for Secondary Ionization"[9] and the stress induced changes in the energy band structure of semiconductors.

II. THEORY

A. Simple Ionization Model

The ionization model to be discussed here for the unstressed junction was introduced by Shockley.[9] The model is empirical in that it involves four-parameters, three of which are adjustable. It includes neither the energy band structure for energies greater than 1 ev from the band edges nor the effective masses of hot carriers. However, the model is found to be in reasonable agreement with experiment.

In the unstressed crystal, the four parameters of the model are as follows:

E_R = Energy of the "Raman" vibration mode.

L_R = Mean-free-path between scattering by "Raman" modes.

E_i = Threshold energy measured from the band edge above which a carrier may produce a hole-electron pair.

L_i = Mean-free-path between ionizations for a carrier with energy greater than E_i.

Using these parameters, an electron with energy greater than E_i generates on the average L_i/L_R phonons per ionization. In the low field limit, which has been shown to be valid for Ge and Si, the average number of scattering events, C, per ionization is

$$C = \frac{L_i}{L_R} \exp(E_i/qL_R F) , \qquad (1)$$

where F is the electric field. The average energy, E_i^*, gained from the field by the carrier per ionization is then

B6

$$E_i^* = \frac{L_i E_R}{L_R} \exp(E_i/qL_R F), \qquad (2)$$

or

$$E_i^* = a \exp(b/F), \qquad (3)$$

where a and b are constants of the material.

The secondary multiplication coefficient, $\alpha(F)$, is

$$\alpha(F) = \frac{qF}{E_i^*} = \frac{qFL_R}{L_i E_R} \exp(-E_i/qL_R F) . \qquad (4)$$

To extend the model to a stressed semiconductor, the effects of multiple conduction and valence levels on the ionization process must be considered.

B. Strain Dependent Ionization Theory

Mechanical strain has the effect of altering the energy band structure of semiconductors. The effects of strain on the conduction and valence band edge points for Ge and Si are reviewed in Appendix A. Under a general strain, the valence band edge not only shifts in energy, but also splits into two levels. The conduction band edge points also shift in energy both relative to the valence levels and to each other. The net results are different energy gaps depending on which valence and conduction band edge points are used to calculate the gap.

In the unstressed crystal, it is not necessary to know the exact dependance of the ionization energy on the energy band structure since this can be determined experimentally. Shockley assumes that for Si, E_i is equal to the band gap.[9] Most investigators,[10,11] however, find a better fit of theory to experimental data using an E_i slightly

larger than the band gap. This is particularly true in Ge. To obtain the numerical results for the effect of stress on breakdown voltage, a model which relates the ionization energy to the energies of the various bands is required. Such a model is discussed in a later section. For the present general development more general assumptions about the ionization process are sufficient.

Consider electron ionization first. In this ionization process, a hot electron in one of the conduction minima (<111> direction minima in Ge and <100> direction minima in Si) excites an electron from one of the valence bands to one of the conduction levels. The result being the creation of an electron-hole pair. The ionization threshold energy is the minimum energy at which the above process can occur. This energy is expected to depend upon the shape of the energy bands in which the ionizing electron and the created pair are located as well as the band edge point energies of the various bands. In the unstressed crystal, all the conduction minima are located at the same energy so there can be only two different threshold energies--one associated with the heavy hole band and one associated with the light hole band. With a stress condition, the conduction levels are split in energy and consequently the threshold energies for the various nondegenerate conduction levels should be different.

Here it is assumed that the ionization threshold energy depends upon the conduction level in which the ionizing hot electron is localed and the valence level in which the created hole is located. This assumes that the created electron is located either in the same

energy minimum as the hot electron or is located in the equivalent
energy minimum in the opposite direction in k-space. This is a
reasonable assumption, since crystal momentum can be more easily
conserved in this process.[12]

An ionization threshold energy E_{nm} can then be associated with
each of the combinations of valence and conduction levels, where n
denotes the conduction level and m denotes the valence level. There
are six such combinations for Si and eight for Ge. Each of the ioni-
zation processes may also have a different mean free path denoted by
L_{nm}.

It will be assumed here that L_R and E_R are not functions of stress,
and that L_R is the same for all the conduction band minima.

Using the above definitions, the number of ionizations which
leaves a hole in the valence level m and an electron in conduction
level n per scattering event, is denoted by $1/C_{nm}$. From Eq. (1) this is

$$\frac{1}{C_{nm}} = \frac{L_R}{L_{nm}} \exp(-E_{nm}/qL_R F) \ . \tag{5}$$

The total number of ionizations per scattering event is obtained by
summing over the ionization processes for a hot electron in a given
conduction minimum and then averaging over the conduction levels. In
averaging over the conduction levels, $1/C_{nm}$ must be multiplied by the
fraction of hot electrons $n_{hn}/\sum\limits_{n} n_{hn}$, in each conduction minimum. The
average number of ionizations per scattering event is then

$$\frac{1}{C} = \frac{\displaystyle\sum_{n=1}^{\beta} n_{hn} \sum_{m=1}^{\gamma} \frac{L_R}{L_{nm}} \exp(-E_{nm}/qL_R F)}{\displaystyle\sum_{n=1}^{\beta} n_{hn}} , \qquad (6)$$

where γ is the number of valence levels and β is the number of conduction levels.

The evaluation of the above formal expression requires an assumption on the distribution of hot electrons among the conduction levels. This distribution of hot electrons depends upon the importance of intervalley and intravalley scattering. If intervalley scattering is negligible, the conduction minima can be considered as independent of each other and

$$n_{hn} \propto \exp(-E_{cn}/kT) , \qquad (7)$$

where E_{cn} is the band edge energy of the conduction level. This is just the equilibrium distribution of electrons among the valleys. On the other hand, if intervalley scattering is predominate, the electrons are more uniformly distributed among the levels because of the increased effective temperature of the hot electrons. On the basis of Shockley's model for ionization, the hot electrons in a given conduction minima are distributed in energy according to the relationship

$$n(E) \propto \exp(-E/qL_R F) . \qquad (8)$$

When intervalley scattering dominates, one would also expect this distribution to hold for electrons among the conduction levels and in this case

$$n_{hn} \propto \exp(-E_{cn}/qL_R F) \ . \hspace{3cm} (9)$$

To keep the assumption more general, the hot electrons are assumed to be distributed among the valleys according to the relationship

$$n_{hn} \propto \exp(-E_{cn}/E_1) \ , \hspace{3cm} (10)$$

where E_1 equals kT and $qL_R F$ for the cases considered above. As is subsequently shown, the low stress case and the very high stress case are independent of the choice of E_1.

The total number of ionizations per scattering event then becomes

$$\frac{1}{C} = \frac{\displaystyle\sum_{n=1}^{\beta} \exp(-E_{cn}/E_1) \ \sum_{m=1}^{\gamma} \frac{L_R}{L_{nm}} \exp(-E_{nm}/qL_R F)}{\displaystyle\sum_{n=1}^{\beta} \exp(-E_{cn}/E_1)} \ . \hspace{1.5cm} (11)$$

The secondary multiplication coefficient for electrons is then given by

$$\alpha(F) = \frac{qFL_R}{E_R} \frac{\displaystyle\sum_{n} \exp(-E_{cn}/E_1) \ \sum_{m} \frac{1}{L_{nm}} \exp(-E_{nm}/qL_R F)}{\displaystyle\sum_{n} \exp(-E_{cn}/E_1)} \ . \hspace{1.2cm} (12)$$

The value of L_{nm} in Eq. (12) may be different for holes created in the lower valence level as compared to holes created in the upper valence level. One would also expect L_{nm} to be slightly dependent upon the direction of the applied field. For example, if the applied field is along a conduction valley direction, one expects a smaller value of L_{nm} for an electron created in the energy minimum along the field direction than for the other energy minimum. The difference between these values depends upon intervalley and intravalley scattering

of hot electrons. If there is sufficient scattering to produce essentially a random distribution of the hot electrons both in and among the conduction valleys, L_{nm} should be essentially the same for all valleys. Sufficient information is not available to permit a detailed evaluation of this parameter. It is also noted that L_{nm} is a linear factor influencing $\alpha(F)$, while the ionization energy E_{nm} is exponentially related to $\alpha(F)$. Thus small changes in E_{nm} tend to be more important in determining $\alpha(F)$ than small changes in L_{nm}. It will therefore be assumed that L_{nm} is the same for all ionization processes. This assumption makes the stress dependence of the secondary multiplication coefficient independent of the direction of the electric field.

For Ge with an electric field in the [100] direction, the component of electric field along each of the valley directions is the same, and the assumption of a constant L_{nm} for all the valleys should be valid. The assumption should be least valid for a field in the ⟨111⟩ directions. For Si, the assumption should be most accurate for a field in the [111] direction and least accurate for a field in the [100] direction.

Assuming that L_{nm} is constant, it can readily be seen from Eq. (12) by considering the unstressed case ($E_{nm} = E_i$) that

$$L_{nm} = \gamma L_i \ . \tag{13}$$

Equation (12) becomes

$$\alpha(F) = \frac{qFL_R}{E_R L_i \gamma} \ \frac{\sum_n \sum_m \exp(-E_{cn}/E_1 - E_{nm}/qL_R F)}{\sum_n \exp(-E_{cn}/E_1)} \ . \tag{14}$$

This is the basic equation used to describe stress dependent voltage breakdown, when electron ionization is the dominant process.

Although the concepts for hole ionization are similar to electron ionization, there are important differences in the details. If the difference in effective mass for the two hole bands is neglected and the ionization energy is assumed to depend only upon the band on which the hot hole exists and upon the conduction minimum to which the election is excited, the evaluation of hole ionization is very similar to that for electron ionization. The difference is that the average number of ionizations per scattering event is obtained by summing over the conduction levels and averaging over the valence levels. This leads to the expression

$$\alpha(F) = \frac{qFL_R}{E_R L_i \beta} \frac{\sum_m \sum_n \exp(E_{vm}/E_1 - E_{nm}/qL_R F)}{\sum_m \exp(E_{vm}/E_1)} , \tag{15}$$

where E_{vm} is the energy of the valence levels.

The large difference in density of states or effective mass for the two valence bands, especially in Ge, makes the neglect of these differences in the above expression open to question. There are, however, factors which make the above expression a better approximation than it would at first appear. First, the average energy of the holes is on the order of E_R which is larger than the thermal energy and, for large energies, the density of states for the two bands become more equal. In Si the density of states are not too different for energies larger than about 0.02 ev.[13] For Ge the energy is considerably larger. A second factor for unequal hole masses is that L_R is different for the

two bands. For a single spherical energy band, L_R has been shown to depend upon the mass as m^{-2}.[14] This partially compensates for the difference in the density of states of the two bands. The compensating effects in the two bands can be thought of in the following manner. The holes in the "light" hole band are more rapidly accelerated than the "heavy" holes to energies sufficient to produce optical phonons or to produce ionizations. Thus, while there are fewer holes in the light hole band, they undergo scatterings and produce ionizations at a faster rate which tends to produce some compensation. Here the purpose is to discuss the major features of stress effects on breakdown so that the simple expression of Eq. (15) is sufficient.

C. Strain Dependent Breakdown

Even when the ionization coefficient is known, the calculation of the breakdown voltage of a p-n junction is difficult, especially when the ionization coefficients for holes and electrons are unequal.[10] Including the effect of stress further complicates the calculation. A simplifying approximation is made to obtain a tractable model. Because of the exponential dependence of α upon $1/F$, the major contributions to α come from the depletion region near the maximum field point. Let α_m be the value of α at the maximum field point, and consider the effect of stress upon α_m. If the maximum junction field is unchanged when stress is applied (i.e. if the junction voltage is unchanged), the ionization coefficient changes and hence the current multiplication factor for the junction changes. On the other hand, if α_m is to remain

constant under stress, the maximum field must change. The approximation which is made here is that at a constant current multiplication factor for the junction, the maximum junction field is changed such that α_m remains constant. The following development is for electron ionization; the corresponding development for hole ionization is similar.

To obtain the change in maximum field required to keep α_m constant, let

$$E_{nm}(\sigma) = E_i + \Delta E_{nm} , \tag{16}$$

and

$$F = F_o + \Delta F , \tag{17}$$

where E_i and F_o are the unstressed values and $E_{nm}(\sigma)$ is the stress dependent ionization energy. Then for electron ionization neglecting second order effects,

$$\alpha_m(F_o + \Delta F) = \frac{qF_o}{E_R} \left(1 + \frac{\Delta F}{F_o}\right) \frac{L_R}{\gamma L_i} \exp[-E_i(1 - \Delta F/F_o)/qL_R F_o] \tag{18}$$

$$\times \frac{\sum_n \sum_m \exp(-\Delta E_{cn}/E_1 - \Delta E_{nm}/qL_R F_o)}{\sum_n \exp(-\Delta E_{cn}/E_1)} .$$

By equating this to the unstressed value of $\alpha_m(F_o)$, it is found that

$$1 = (1 + \frac{\Delta F}{F_o}) \frac{1}{\gamma} \exp(E_i \Delta F/qL_R F_o^2) \frac{\sum_n \sum_m \exp(-\Delta E_{cn}/E_1 - \Delta E_{nm}/qL_R F_o)}{\sum_n \exp(-\Delta E_{cn}/E_1)} . \tag{19}$$

This expression relates the change in the maximum junction electric field to the stress induced changes in the energy levels.

To obtain the change in junction voltage at a constant current multiplication factor, the change in maximum junction field must be

related to the junction voltage. The maximum field is proportional to $V^{1/2}$ for a step junction.[15] Thus for small changes

$$\frac{\Delta V}{V_B} \approx 2 \frac{\Delta F}{F_o} , \qquad (20)$$

where V_B is the unstressed breakdown voltage at a constant current level. For most p-n junctions, the step junction approximation is reasonably accurate at large reverse bias.

For small changes in the breakdown voltage ($\Delta V/V_B \ll 1$), Eq. (19) can be solved for the change in breakdown voltage to give

$$\frac{\Delta V}{V_B} = - \frac{2}{(1 + E_i/qL_RF_o)} \, \ell n \, \{\frac{1}{\gamma} \frac{\sum_n \sum_m \exp(-\Delta E_{cn}/E_1 - \Delta E_{nm}/qL_RF_o)}{\sum_n \exp(-\Delta E_{cn}/E_1)}\} . \qquad (21)$$

For hole ionization a similar development leads to

$$\frac{\Delta V}{V_B} = - \frac{2}{(1 + E_i/qL_RF_o)} \, \ell n \, \{\frac{1}{\beta} \frac{\sum_n \sum_m \exp(\Delta E_{vm}/E_1 - \Delta E_{nm}/qL_RF_o)}{\sum_m \exp(\Delta E_{vm}/E_1)}\} . \qquad (22)$$

In the low stress region ($\Delta E_{nm}/qL_RF_o \ll 1$ and $E_{cn}/E_1 \ll 1$), the equations for both electron and hole ionization simplify to

$$\frac{\Delta V}{V_B} = \frac{2}{qL_RF_o(1 + E_i/qL_RF_o)} \, \frac{1}{\gamma\beta} \sum_{n,m} \Delta E_{nm}. \qquad (23)$$

In this limit, the factor $(1/\gamma\beta) \sum \Delta E_{nm}$ is the simple average of the changes in the ionization threshold energies.

To carry the development further and obtain numerical results requires a model for the changes in the ionization threshold energies with stress. One assumption is that the ionization threshold energy

is equal to the band gap. This implies that momentum is supplied by phonons. Neglecting phonon assisted processes the ionizing electron must have both enough energy to create the hole-electron pair (E_g) and also enough excess energy to conserve both momentum and energy between the single incident particle and the three final particles. For a direct band gap material the minimum energy is readily found to be $3E_g/2$. For an indirect band gap material such as Ge and Si, it has recently been shown[12] that a consideration of energy and crystal momentum conservation leads to an expression of the form

$$E_i \simeq k_1 + k_2 E_g, \qquad (24)$$

where k_1 and k_2 are constants depending in rather complex ways on the effective masses of the three particles. The effective masses entering into k_1 and k_2 are not the band edge effective masses, but are the masses which give the best fit to the energy bands over the energy range from zero to the ionization energy. Based upon this model, the ionization threshold energy between a conduction level and a valence level is

$$E_{nm} = k_1 + k_2(E_{cn} - E_{vm}). \qquad (25)$$

This is the model used in this work to relate the band structure to the ionization energy.

Stress can have two effects on E_{nm} as defined above through changes in the energy levels ($E_{cn} - E_{vm}$) and through changes in the shape (or effective mass) of the energy levels which results in changes in k_1 and k_2. The effect of stress on the band edge energy levels is

known but the effect of stress on the effective mass of the bands especially at large energies has not been as thoroughly investigated. Some work has been done on the effective masses near the band edges, but the effective mass at large energies probably changes less than near the band edges, especially for the hole bands. Near the band edges there is no first order change in the effective mass with stress for electrons.[16,17] Also under compression stress in the [100] and [111] directions there is no first order change in the effective mass of the lower valence level.[18] Based upon the above facts, k_1 and k_2 have been taken as independent of stress for our first order model and the changes in ionization energy are taken to be

$$\Delta E_{nm} = k_2(\Delta E_{cn} - \Delta E_{vm}). \tag{26}$$

For equal hole and electron masses, k_2 is 0.59 for electron ionization in Ge and 0.63 for hole ionization in Ge and for both electron and hole ionization in Si.[12]

Using the above model for ionization, the change in breakdown voltage at low stress is given from Eq. 24 as

$$\frac{\Delta V}{V_B} = \frac{2k_2 KP}{E_i} \left(1 + \frac{qL_R F_o}{E_i}\right)^{-1}, \tag{27}$$

where K is the coefficient relating the change in band gap to hydrostatic pressure and P is the hydrostatic component of stress.

Typical values of the parameters in Eq. (27) are listed in Table I. Using these values, Eq. (27) reduces to

TABLE I: Typical Values of K, F_o, L_R and E_i in Ge and Si

K(10^{-12}ev cm^2/dyne)	F_o(10^5 volts/cm)	L_R(A°)	E_i(ev)
Si -1.5[a]	5[b]	$50-70$[d]	1.3[f,g]
Ge 5.0[e]	0.82[c]	65	1.0[f], 0.93[g]

a. W. Paul and G. L. Pearson, Phys. Rev., 98, 1755 (1955).

b. Ref. 9

c. R. Yee, J. Murphy, A. D. Kurtz, and H. Bernstein, J. Appl. Phys. 30, 596 (1959).

d. Ref. 10

e. P. W. Bridgman, Proc. Am. Acad. Arts Sci. 79, 129 (1951).

f. Values calculated for electron ionization, Ref. 12.

g. Values calculated for hole ionization, Ref. 12.

$$\text{Si:} \quad \frac{\Delta V}{V_B} = -1.2 \times 10^{-12} P \ cm^2/dyne, \tag{28}$$

$$\text{Ge:} \quad \frac{\Delta V}{V_B} = 6.0 \times 10^{-12} P \ cm^2/dyne. \tag{29}$$

As can be seen from Eq. (27), at low stress levels, $\Delta V/V_B$ is independent of stress orientation since only the hydrostatic component of stress remains in the expressions. It is significant that at low stress levels, $\Delta V/V_B$ has an opposite sign for Ge and Si.

Returning to Eq. (22), it is seen that in the high stress case, the minimum conduction level and the maximum valence level will predominate, i.e., the ΔE_{nm} with the largest negative value. The corresponding ΔE_{cn} is negative. In general for uniaxial compression stresses at least one of the ΔE_{nm}'s will be negative for both Ge and Si. Therefore, for high stresses, $\Delta V/V_B$ decreases for both materials, and is independent of the value of E_1. This is true for both electron and hole ionization.

For a given value of the ratio $E_1/qL_R F_o$, the change in breakdown voltage under stress can be calculated from Eqs. (21) and (22) for electron and hole ionization respectively. To account for the increased effective temperature of the carriers in the high field, the value $E_1 = qL_R F_o$ has been used in most of the calculations. The following deformation potentials were used in evaluating the shifts in the energy levels (see Appendix A):

Ge	Si

$$D_u = 3.15 \qquad\qquad D_u = 2.04$$

$$D_u' = 6.06 \qquad\qquad D_u' = 2.68$$

$$\Xi_u = 19.2 \qquad\qquad \Xi_u = 11$$

$$D_d - (\Xi_d + \Xi_u/3) = 4.82 \qquad\qquad D_d - (\Xi_d + \Xi_u/3) = -1.44$$

The calculated changes in breakdown voltage in Ge with stress for electron and hole ionization are shown in Figs. 1 and 2 for uniaxial stresses in the [100], [111], and [110] directions. The curves are plotted in terms of a normalized stress and a normalized voltage change given by the expressions

$$\sigma' = (E_g/qL_R F_o)\sigma,$$

(30)

$$\frac{\Delta V'}{V_B} = \frac{1}{2}\left(1 + \frac{E_i}{qL_R F_o}\right)\frac{\Delta V}{V_B}.$$

The curves illustrate the increase in breakdown voltage in Ge at low stresses, independent of the stress orientation, and the decrease in breakdown voltage at large stress values. It is also noted that for electron ionization, the changes for the [100] direction require approximately an order of magnitude larger stress than the other two directions.

A comparison of the theoretical changes (for electron ionization) in breakdown voltage for two values of E_i are shown in Fig. 3. The two values of E_i are kT and $qL_R F_o$. It is recalled that the value kT

corresponds to completely isolated conduction minima with no intervalley scattering. The two cases are seen to have the same general behavior for the two values of qL_RF_o shown.

Similar curves to those in Figs. 1 and 2 can be obtained for Si. In this case, however, there is only a decrease in breakdown voltage at both low and high stress. It can also be seen from Table I, that the normalization factor (E_g/qL_RF_o) has the typical value of 12.2 for Ge and 3.4 for Si. This means that the low stress approximation is valid to higher stress levels in Si than in Ge. In fact for practical stress values (values up to 10^{10} dynes/cm^2) the low stress approximation is valid.

III. COMPARISON WITH EXPERIMENT

The theory developed in the preceeding sections gives the change in the junction breakdown voltage as a function of stress for a constant multiplication factor. When making measurements on practical devices, one measures the applied voltage which differs from the junction voltage by the built-in potential. The effect of stress on the built-in potential is discussed in Appendix B. It is shown there that if V_B is large (>10 volts), then the changes in the built-in potential can be neglected compared to changes in the junction voltage.

When comparing experiment to the present theory, one must keep in mind that the voltage must be measured at a constant multiplication factor. The voltages cannot be measured at a constant current level unless the constant current level is several orders of magnitude above

the saturation current. The reason this is necessary is that the saturation current is also stress sensitive. The multiplication factor, M is related approximately to the applied voltage, V_A, by[19]

$$M = \frac{1}{1 - (V_A/V_B)^n} , \qquad (31)$$

where n is a constant for the device. The diode current, I, is given by

$$I = I_s M , \qquad (32)$$

where I_s is the reverse saturation current. Combining Eqs. (30) and (31) gives

$$V_A = V_B(1 - I_s/I)^{1/n} . \qquad (33)$$

If $M \gg 1$,

$$V_A = V_B(1 - \frac{1}{n}\frac{I_s}{I}) . \qquad (34)$$

Neglecting second order terms, the relative change in the applied voltage at constant current is

$$\frac{\Delta V_A}{V_B} \approx \frac{\Delta V}{V_B} - \frac{\Delta I_s}{nI} . \qquad (35)$$

The change in applied junction voltage is then less than the change in the breakdown voltage by the factor $V_B\Delta I_s/nI$. This is negligible only if nI is large compared with ΔI_s. Using Eq. (35), $\Delta V/V_B$ can be evaluated from an experimental plot of current versus voltage as a function of stress if ΔI_s is known as a function of stress. For a uniformly stressed junction in which the total junction area is stressed, $\Delta I_s/I_{so}$ is given by[6]

$$\frac{\Delta I_s}{I_{so}} = \frac{1}{\gamma\beta} \ \Sigma \ \exp(\frac{\Delta E_{cn} - \Delta E_{vm}}{kT}) - 1 = (f(\sigma) - 1) \ , \tag{36}$$

where I_{so} is the unstressed saturation current. This gives the following expression for $\Delta V_A/V_B$

$$\frac{\Delta V_A}{V_B} \approx \frac{\Delta V}{V_B} - \frac{(f(\sigma) - 1)}{nM_o} \ , \tag{37}$$

where M_o is the multiplication factor between the unstressed saturation current and the current at which V_A is measured.

In Si, it has been found experimentally that uniaxial compression stress causes a linear decrease in $\Delta V_A/V_B$ by the factor[7,8]

$$\frac{\Delta V_A}{V_B} \approx (-1 \times 10^{-12} cm^2/dyne) \ \sigma \ , \tag{38}$$

where σ is the magnitude of the uniaxial stress. The theoretical proportionality constant is $-0.4 \times 10^{-12} cm^2/dyne$ from Eq. (28) where $\sigma = P/3$. The theoretical value was calculated using the value $-1.5 \times 10^{-12} ev \ cm^2/dyne$ for the hydrostatic pressure dependance of the band gap. Although this value is normally used, there is some variation in it depending on the method of measurement and the stress level.[20] In fact, it has been shown that the coefficient is stress dependant and at high stress levels, it is approximately $-2.5 \times 10^{-12} ev \ cm^2/dyne$.[21] Using the latter value gives a proportionality constant in Eq. (28) of approximately $-0.67 \times 10^{-12} cm^2/dyne$.

As predicted by the theory, very little if any variation in the phenomenon with crystal orientation is seen experimentally in Si.[7,8]

Preliminary measurements on breakdown voltage changes with hydrostatic

pressure have given results which were an order of magnitude smaller

than those for uniaxial stress.[8] The present theory predicts that

the voltage changes induced by hydrostatic pressure should be three

times those of uniaxial stresses. At the present time, there is no

explanation for these results.

Rindner has made measurements of the breakdown voltage change

as a function of uniaxial stress in Ge diodes.[7] The experimental

points in Fig. 4 are the results he obtained from a [110] and a [100]

uniaxial stress. The solid curves in the figure were calculated using

the theory for hole ionization. The hole ionization curves were

used because the secondary multiplication factor for holes is larger

in Ge. The normalizing factor $E_g/qL_R F_o$ was considered as an adjustable

parameter in fitting the theory to experiment. The values needed to

fit the experimental data are approximately 8.5 and 16 for the [110]

and [100] directions respectively. These values are in good agreement

with the value of 12.6 calculated from the data in Table I. Attempts

to fit the theory to published experimental data taken for [111] oriented

uniaxial stress[7] have been less successful. It should be noted that

the data can also be made to agree with the theory for electron ioniza-

tion by a suitable choice of $L_R F_o$. For the [100] direction, however,

an unreasonably small value of $L_R F_o$ is required to produce a good fit

to the data. It is difficult to draw too many conclusions from these

comparisons without knowing the repeatability of the experimental

results and the conditions under which the data was obtained. As

shown in Fig. 4, the experimental results and theoretical curves deviate from each other for low stress levels. This could result from nonuniform breakdown in the junction due to microplasma and surface effects.

V. DISCUSSION

The theoretical model for stress induced changes in breakdown voltage is in generally good agreement with experiment. In view of the simple model used for breakdown and the assumptions necessary to arrive at an analytical expression for the effect, it is surprising that the theory agrees as well as it does. The fact that the theory predicts an orientation dependance in Ge but no such dependance in Si in agreement with the experimental observations, lends considerable support to the theory.

Care must be exercised when comparing the theory to experiment since most practical devices do not exhibit uniform breakdown. Also, it is difficult to determine what percentage of the multiplication factor results from hole ionization as compared to electron ionization. These factors are necessary to make accurate comparison between theory and experiment. The comparison for Ge made in the preceeding section was based on the assumption that the ionization energy is that of holes.

APPENDIX A

Both Ge and Si have multiple conduction minima in k-space. Ge has eight such minima which lie in the <111> direction and are located at the L_1 point in k-space while silicon has six conduction minima which occur in the <100> direction and are located approximately 85% of the distance from k = (000) to the X_1 symmetry point. The maximum valence levels, Γ'_{25}, for both Ge and Si are located at k = (000). The Γ'_{25} level is degenerate in energy. If the crystal is mechanically deformed, the crystal symmetry and the lattice spacings are altered and hence the energy bands change.

Herring and Vogt have considered the effect of mechanical strain on the conduction minima in both Ge and Si.[22] The results of their work are summarized in Table II. The valley directions are identified by the subscripts on the conduction energy level changes ΔE_c's. The Ξ's are deformation potential constants and the e's are conventional strains. Table III lists some of the values of the deformation potential constants. The values appearing in brackets are theoretical values calculated by Kleinman, et al.[23-25]

The effects of mechanical strain on the valence levels of Ge and Si are much more complicated than the effects on the conduction levels. Kleiner and Roth have considered the effects of strain on the Hamiltonian of the valence band edge.[26] Diagonalizing their expression for the Hamiltonian gives the following expression for the change in energy of the valence level[27]

$$\Delta E = D_d e \pm [(\tfrac{2}{3} D_u)^2 (e_1^2 + e_2^2 + e_3^2 - e_1 e_2 - e_1 e_3 - e_2 e_3)$$

$$+ \tfrac{1}{3} (D_u')^2 (e_4^2 + e_5^2 + e_6^2)]^{1/2}. \qquad (A1)$$

The D's appearing in the above expression are the valence band deformation potential constants. From Eq. (A1), it is seen that there is not only a shift of the level due to D_d, but also a splitting of the level due to D_u and D_u', which removes the degeneracy. For convenience of notation, let the upper band, positive sign of Eq. (A1), be ΔE_{V1} and the lower band, negative sign, be ΔE_{V2}.

TABLE II. Equations describing the change in the band edge points in the conduction band of Ge and Si as a function os strain.

Valley Direction	Band Edge	Energy
Ge		
$[111]$, $[\bar{1}\bar{1}\bar{1}]$	ΔE_{c1}	$(\Xi_d + \Xi_u/3)e + \Xi_u(e_4 + e_5 + e_6)/6$
$[1\bar{1}\bar{1}]$, $[\bar{1}11]$	ΔE_{c2}	$(\Xi_d + \Xi_u/3)e + \Xi_u(e_4 - e_5 - e_6)/6$
$[1\bar{1}1]$, $[\bar{1}1\bar{1}]$	ΔE_{c3}	$(\Xi_d + \Xi_u/3)e + \Xi_u(-e_4 + e_5 - e_6)/6$
$[11\bar{1}]$, $[\bar{1}\bar{1}1]$	ΔE_{c4}	$(\Xi_d + \Xi_u/3)e + \Xi_u(-e_4 - e_5 + e_6)/6$
Si		
$[100]$, $[\bar{1}00]$	ΔE_{c1}	$\Xi_d e + \Xi_u e_1$
$[010]$, $[0\bar{1}0]$	ΔE_{c2}	$\Xi_d e + \Xi_u e_2$
$[001]$, $[00\bar{1}]$	ΔE_{c3}	$\Xi_d e + \Xi_u e_3$

TABLE III. Deformation Potential Coefficients (ev/unit dilation) for Ge and Si. (Kleinman's theoretical values are shown in brackets.)

Coefficient	Si	Ge
D_d	[-2.09]	[-2.09]
D_u	2.04[a], [3.74]	3.15[b], [3.74]
D'_u	2.68[a], [4.23]	6.06[b], [3.6]
Ξ_d	[-4.99]	[-10.16]
Ξ_u	11[c], 8.3[d], [+9.6]	19.2[e], [11.4]
Ξ'_u	5.7[f]	
$D_d - (\Xi_d + \frac{1}{3} \Xi_u)$	-1.44[g], [-0.30]	4.82[g], [4.27]

a. J. C. Hensel and G. Feher, Phys. Rev. 129, 1041 (1963).

b. J. J. Hall, Phys. Rev. 128, 68 (1962).

c. D. K. Wilson and G. Feher, Phys. Rev. 124, 1968 (1961).

d. J. E. Aubrey, W. Gubler, T. Henningsen and S. H. Koenig, Phys. Rev. 130, 1667 (1963).

e. H. Fritzche, Phys. Rev.

f. J. C. Hensel and H. Hasegawa, paper presented at the International Conference on the Physics of Semiconductors, Paris, July 1964, and private communications.

g. W. Paul, J. Phys. Chem. Solids 8, 196 (1959).

APPENDIX B: Effect of Stress on V_o.

The applied voltage differs from the junction voltage in reversed biased junctions by the built-in potential, i.e.

$$V = V_A + V_o , \tag{B1}$$

where V is the junction voltage, V_A is the applied voltage, and V_o is the built-in voltage. Mechanical stress not only changes the junction voltage, but it also can change the built-in voltage. The relative change in the applied voltage at breakdown is

$$\frac{\Delta V_A}{V_B} = \frac{\Delta V}{V_B} - \frac{\Delta V_o}{V_B} . \tag{B2}$$

The built-in voltage for a step junction is

$$V_o = \frac{kT}{q} \ln \left(\frac{N_A N_D}{n_i^2}\right) , \tag{B3}$$

when N_A and N_D are the acceptor and donar concentrations on the p and n side of the junction respectively and n_i is the intrinsic carrier concentration ($n_i^2 = pn$).

Wortman, Hauser, and Burger have treated the effect of stress on p and n.[6] Using their results and forming the pn product gives

$$n_i^2 = pn = p_o n_o f(\sigma) , \tag{B4}$$

when $p_o n_o$ is the unstressed intrinsic carrier concentration. The factor $f(\sigma)$ is given by

$$f(\sigma) = \frac{1}{\gamma\beta} \Sigma \exp\left(\frac{\Delta E_{cn} - \Delta E_{vm}}{kT}\right) . \tag{B5}$$

The change in V_o divided by V_B becomes

$$\frac{\Delta V_o}{V_B} = - \frac{kT}{qV_B} \ln f(\sigma). \tag{B6}$$

At room temperature $qL_R F_o / E_g$ is approximately equal to kT for Ge so that

$$\frac{\Delta V_o}{V_B} \approx 4 \frac{E_i}{qV_B} \frac{\Delta V}{V_B} . \tag{B7}$$

Therefore, for Ge, if V_B is larger than several volts, $\Delta V_o / V_B$ can be neglected compared to $\Delta V / V_B$.

In Si, $qL_R F_o$ is approximately 10 kT so that, except for hydrostatic stresses, the exponential terms will contribute to the built-in potential at a lower stress level than they do for the junction voltage. Therefore, if V_B is not large, the change in the built-in potential can reduce the change in the applied voltage. In fact, if V_B is on the order of one volt, the built-in potential change can dominate the applied voltage change.

Footnotes and References

* This work was supported in part by the National Aeronautics and Space Administration under Contract No. NASr-222.

[1] H. Hall, J. Bardeen, and G. Pearson, Phys. Rev. **84**, 129 (1951).

[2] M. E. Sikorski, Physical Acoustics, Vol. 1, Part B, Edited by W. P. Mason (Academic Press, New York, 1964), Cha. 12.

[3] W. Rindner, J. Appl. Phys. **36**, 2513 (1965).

[4] T. Imai, M. Uchida, H. Sato, and A. Kobayashi, Japanese J. Appl. Phys., **4**, 102 (1965).

[5] Y. Matukura, Japanese J. Appl. Phys. **3**, 256 (1964).

[6] J. J. Wortman, J. R. Hauser and R. M. Burger, J. Appl. Phys. **35**, 1222 (1964).

[7] W. Rindner, Appl. Phys. Letters **6**, 225 (1965).

[8] A. Goetzberger and R. H. Finch, J. Appl. Phys. **35**, 1851 (1964).

[9] W. Shockley, Solid State Electronics **2**, 35 (1961).

[10] C. A. Lee, R. A. Logan, J. J. Kleimack, and W. Wiegmann, Phys. Rev. **134**, A761 (1964).

[11] A. G. Chynoweth and K. G. McKay, Phys. Rev. **168**, 29 (1957).

[12] J. R. Hauser, J. Appl. Phys. **37**, 507 (1966).

[13] E. O. Kane, J. Phys. Chem. Solids, **1**, 82 (1956).

[14] W. Shockley, Bell System Techn. J. **30**, 990 (1951).

[15] W. Shockley, Bell System Techn. J. **28**, 435 (1949).

[16] J. Hensel, H. Hasegawa and M. Nakayama, Phys. Rev. **138**, A225 (1965).

Footnotes and References (continued)

[17] C. Herring, Bell System Tech. J. 34, 237 (1955).

[18] H. Hasegawa, Phys. Rev. 129, 1029 (1963).

[19] S. L. Miller, Phys. Rev. 99, 1234 (1955).

[20] W. Paul and D. M. Warschauer, Solids Under Pressure (McGraw-Hill Book Company, Inc., New York), 1963, Chap. 8.

[21] M. I. Nathan and W. Paul, Phys. Rev. 128, 38 (1962).

[22] C. Herring and E. Vogt, Phys. Rev. 101, 944 (1956).

[23] L. Kleinman, Phys. Rev. 128, 2614 (1962).

[24] L. Kleinman, Phys. Rev. 130, 2283 (1963).

[25] I. Goroff and L. Kleinman, Phys. Rev. 132, 1080 (1963).

[26] W. H. Kleiner and L. M. Roth, Phys. Rev. Letters 2, 334 (1959).

[27] J. J. Wortman, Ph.D. Dissertation, Duke University, (1965).
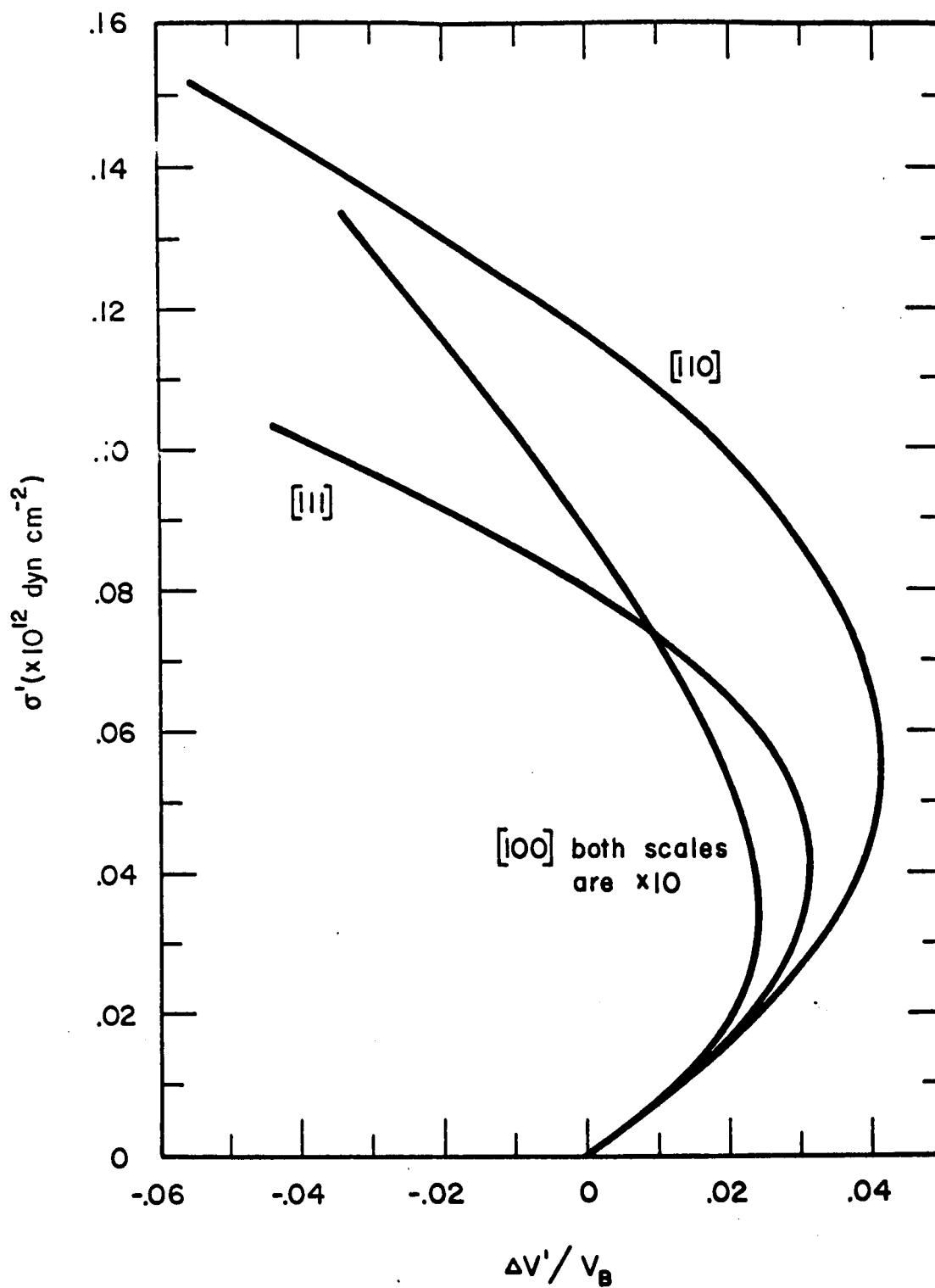
Figure 1.   Theoretical values of $\Delta V'/V_B$ as a function of σ' for electrons
with stress in the [100], [011], [111] directions in germanium.
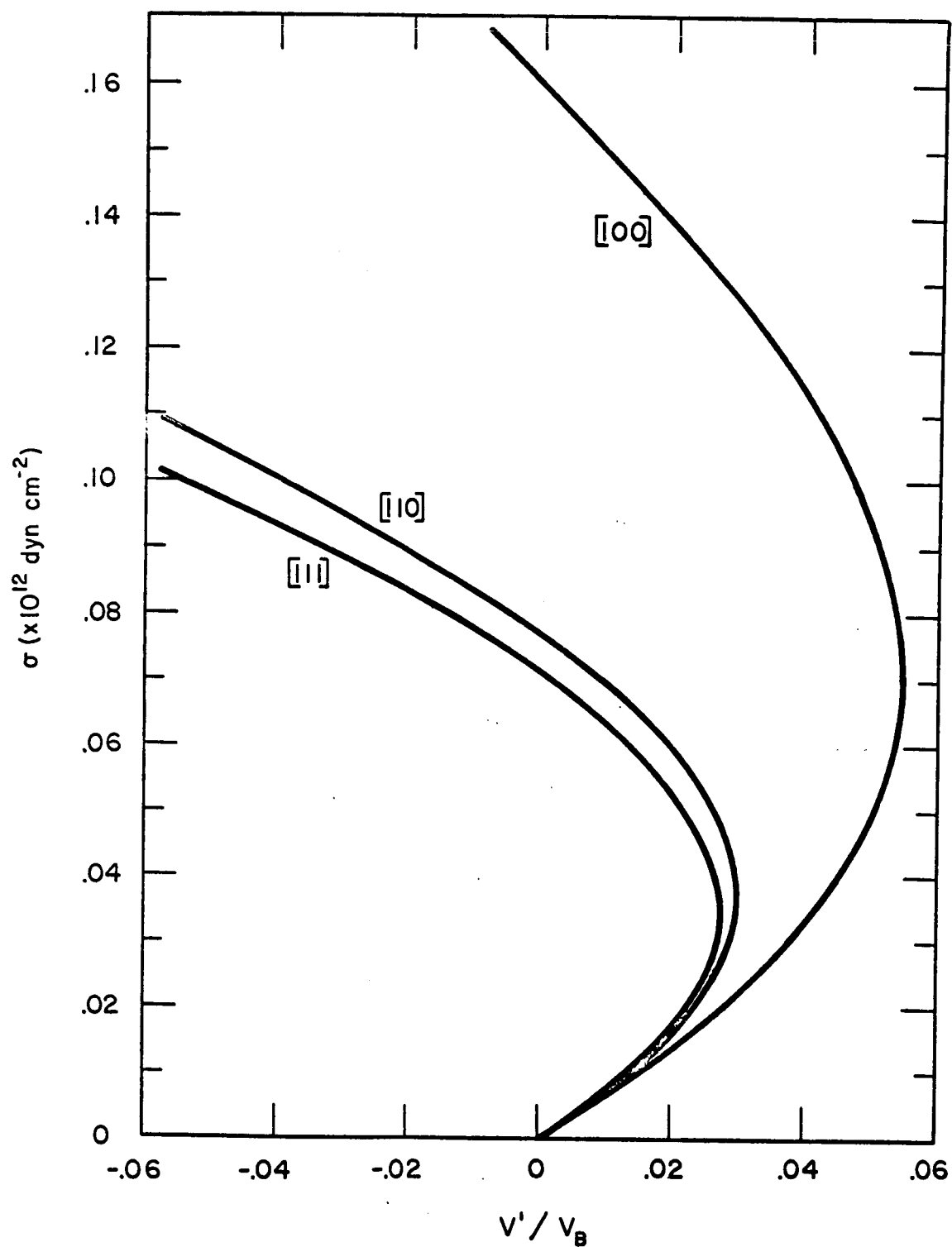
Figure 2. Theoretical values of $\Delta V'/V_B$ as a function of σ' for holes with stress in the [100], [011], [111] directions in germanium.

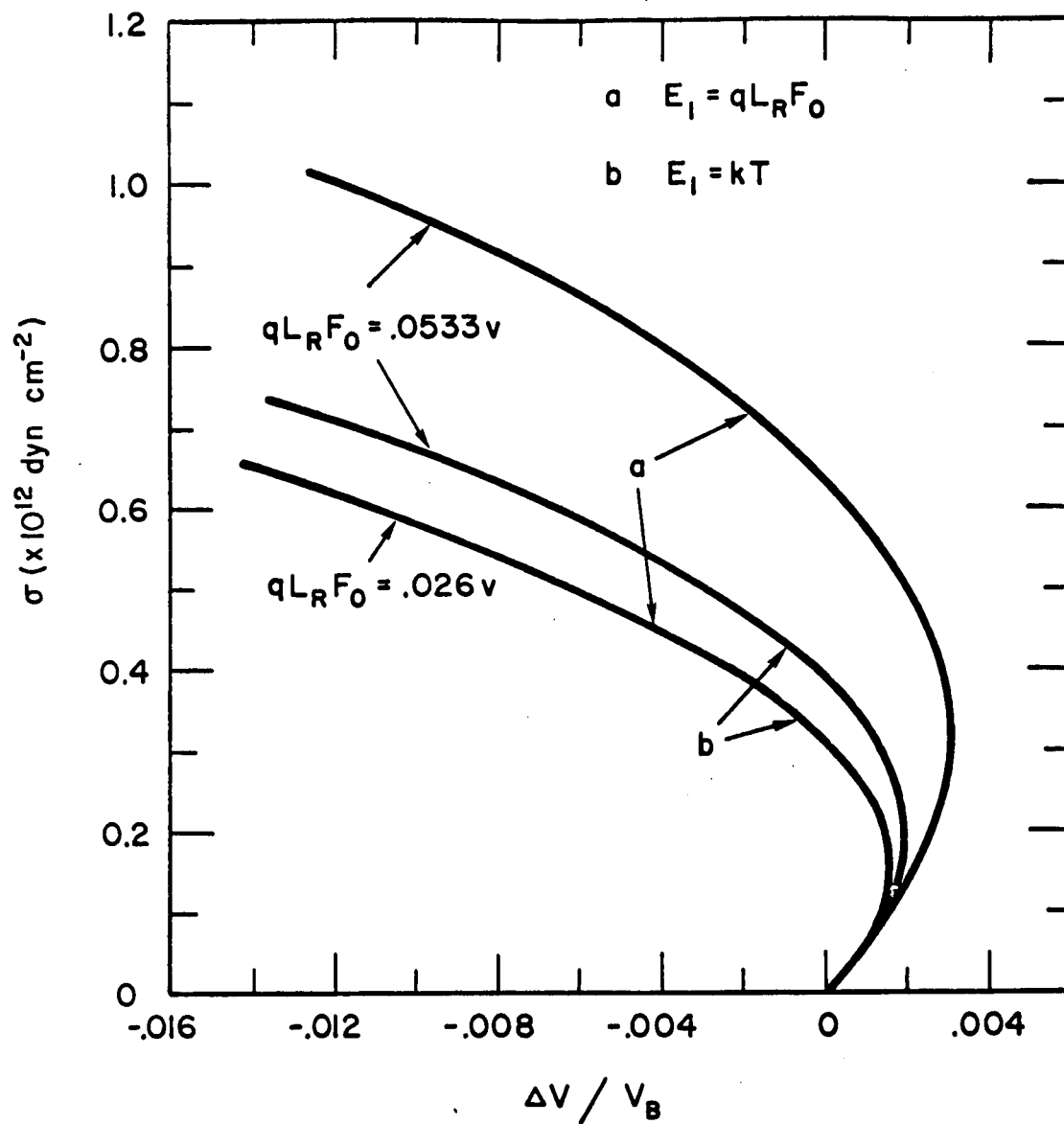Figure 3. Comparison of electron breakdown in germanium for two values
of E_1 (stress in [111] direction).
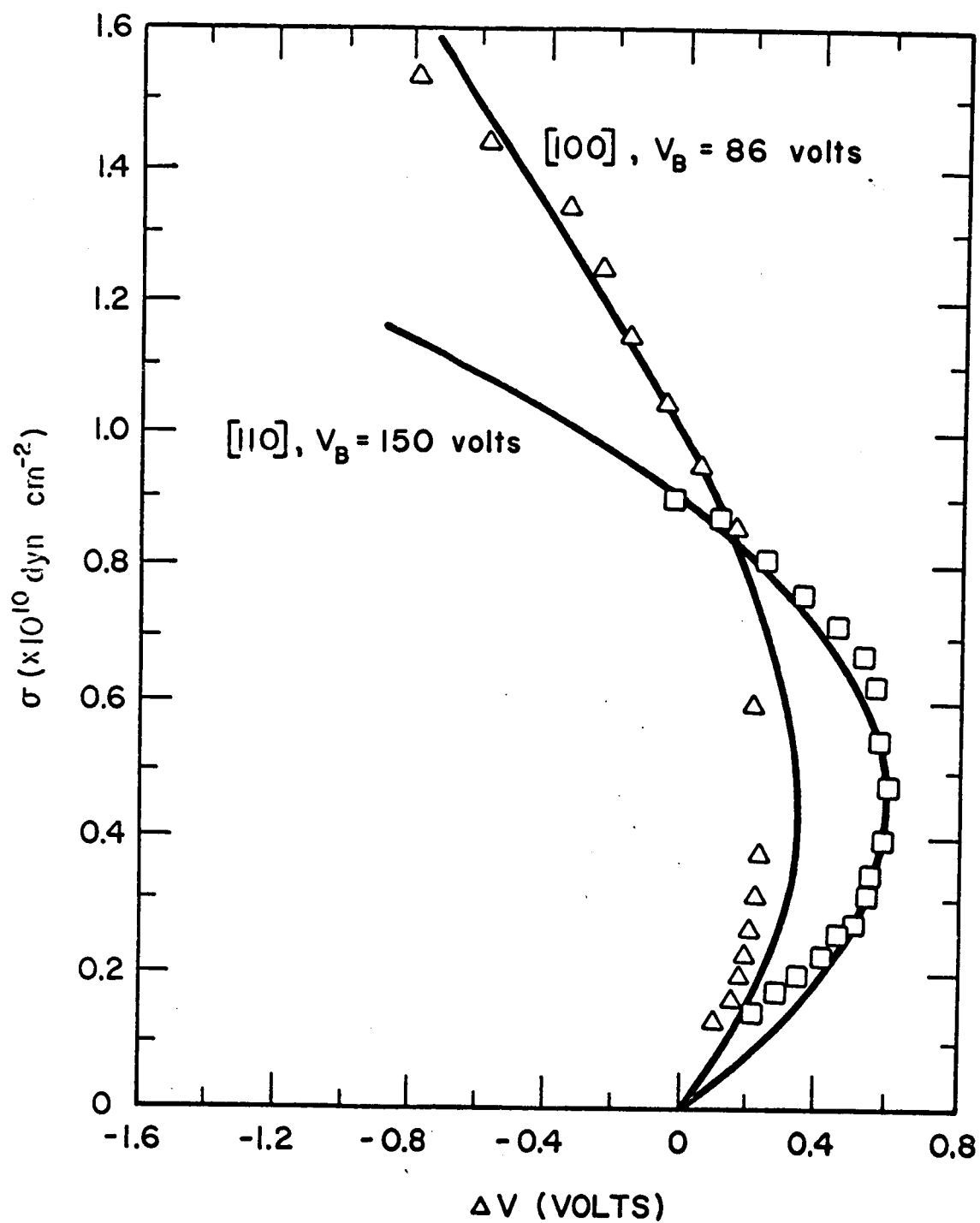
Figure 4. Change in breakdown voltage as a function of stress in germanium diodes. The solid lines are calculated values and the data points are experimental values reported by Rindner.[7]

Appendix C

EFFECT OF MECHANICAL STRESS ON p-n JUNCTION DEVICE CHARACTERISTICS II.
GENERATION-RECOMBINATION CURRENT

Appendix C

Effect of Mechanical Stress on p-n Junction Device Characteristics II.

Generation-Recombination Current

J. J. Wortman and J. R. Hauser
Research Triangle Institute, Durham, North Carolina

Effect of Mechanical Stress on p-n Junction Device Characteristics II.

Generation-Recombination Current[*]

J. J. Wortman and J. R. Hauser

Research Triangle Institute, Durham, North Carolina

A theoretical model is developed for the effects of stress on generation-recombination currents in p-n junctions. The model is based on stress induced changes in the energy band structure of the semiconductor and energy changes of the generation-recombination centers. Experimental data taken on silicon mesa diodes are presented which is in good agreement with the theory.

I. INTRODUCTION

Two basic models have been postulated which attempt to explain stress induced changes in the current-voltage characteristics of p-n junctions. The first of the models suggests the existance of reversible stress induced generation-recombination centers.[1,2] The second model proposes that stress induced changes in the energy band structure of the material is responsible for the observed phenomenon. The latter model is discussed in the first paper[3] (hereafter called I) for the case of the "ideal" or Shockley current in p-n junctions. The analysis presented in I has been found to agree well with experimental results provided the junction current is of the "ideal" type.[4]

In silicon diodes, the junction current is often dominated by generation-recombination currents. This is particularly true for reverse biases and low forward biases. The presence of only generation-recombination currents is some experimental devices and the presence of both the "ideal" and generation-recombination currents in other devices complicate the explanation of the observed effects of stress on the I-V characteristics. A further complication arises when a device is stressed to the point where dislocations (non-reversible) are introduced.

The bulk generation-recombination current in p-n junctions is determined by the density of trapping centers, the trapping cross section and their energy level relative to the conduction and valence levels. It is the purpose of this paper to show that generation-recombination current is altered by stress through changes induced in the energy band structure of the semiconductor and the energy of the trap levels. Many

of the observations which have been attributed to the creation of reversible trap centers are explained by the model which is presented in the following discussion.

II. MODEL

The Shockley-Read-Hall theory[5,6] for the generation-recombination process of single energy trap-centers is used. An expression has been given by Sah, et al.,[7] for the current resulting from generation-recombination centers in a p-n junction. This expression must be solved by numerical means. One of the authors (JRH)[8] has developed an approximate expression for the current which is easier to manipulate and is adequate for the present purposes. This approximation is

$$I_r = \frac{q n_i^2 \, WA \, [\exp(qV/kT) - 1]}{(\tau_{po} n_1 + \tau_{no} p_1) + (\tau_{po} + \tau_{no}) n_i \, \exp(qV/2kT)} \, , \qquad (1)$$

where W is junction width, A is junction area, τ_{no} (τ_{po}) is the lifetime of electrons (holes) injected into highly p-type (n-type) material, $p_1(n_1)$ is the density of holes in the valence (conduction) band when the Fermi energy falls at the trap energy.

The parameters in Eq. (1) which are subject to change with stress are n_i, τ_{po}, τ_{no}, p_1, n_1, and W. For the present purposes, it will be assumed that τ_{po} and τ_{no} are not functions of stress—at least in the range where reversible stress induced current changes are observed. It is well known that large stress can introduce dislocation which may result in trap centers and consequently decreased lifetimes. It is likely, however, at room temperature that most of these centers permanently

remain in the crystal or if they change, they have a long time constant. Implied in the assumption that τ_{no} and τ_{po} are constant is the assumption that the capture cross section is stress independent.

The effect of stress on the intrinsic carrier concentration, n_i, can be determined by considering the effects of stress on the energy band structure. Since $n_i^2 = pn$, it is seen by forming this product from Eqs. (4) and (6) of I that $n_i^2 = n_{io}^2 \gamma_\nu$. There γ_ν [Eq. (9) of I] gives the stress dependance and n_{io} is the unstressed intrinsic carrier concentration. At high stress levels γ_ν is related to stress as follows:

$$\gamma_\nu = C_1 \exp C_2 \sigma , \qquad (2)$$

where C_1 and C_2 depend on the material and stress orientation and σ is the magnitude of the applied stress.

The width of the space charge region, W, is a function of the junction voltage. For a step junction, the width is

$$W \propto (V_o - V)^{1/2} , \qquad (3)$$

where V_o is the built-in junction potential. For a step junction, V_o is related to the intrinsic carrier density by

$$V_o = \frac{kT}{q} \ln\left(N_a N_d / n_i^2\right) , \qquad (4)$$

where $N_a (N_d)$ is the acceptor (donor) density on the p(n) side of the junction. Since V_o is logarithmically related to changes in n_i^2, the influence of changes in V_o on W and hence the current will be small in comparison with changes in n_i.

This leaves only p_1 and n_1 to be discussed. These are given by[5]

$$P_1 = n_i \exp(E_t - E_i)/kT \tag{5}$$

and

$$n_1 = n_i \exp(E_i - E_t)/kT , \tag{6}$$

where E_i is the intrinsic Fermi energy and E_t is the trap energy. Unfortunately, there is little information available on the effect of a general stress on the trap energies of the many types of generation-recombination centers. Stress induced changes in the trap energy would, of course, be a function of the type of trap and its energy with respect to the conduction and valence levels. Data has been reported on the effects of hydrostatic stress on such impurities as gold.[9] A study of the effects of stress on the lifetime in bulk semiconductors with known trap centers could lead to useful information concerning the changes in p_1 and n_1. In the absence of this information, it is not possible to predict the dependence of I_r on the stress except in the large forward biased case where p_1 and n_1 cancel out of the expression.

In the large forward bias case where

$$(qV/2kT) \gg \ln \frac{\tau_{po}n_1 + \tau_{no}p_1}{n_i(\tau_{po} + \tau_{no})} ,$$

Eq. (1) reduces to

$$I_r = \frac{qn_i WA}{\tau_{no} + \tau_{po}} \exp(qV/2kT) . \tag{7}$$

Based on the foregoing assumptions and assuming W is approximately constant (good for $V < V_o$), n_i is the only stress dependant parameter. This gives

$$I_r = I_{or} \sqrt{\gamma_\nu} \exp(qV/2kT) \ , \tag{8}$$

where I_{or} is the unstressed generation-recombination current. If only part of the junction area is stressed, A_s, then

$$I_r = \frac{A_t - A_s}{A_t} + \frac{A_s}{A_t} \sqrt{\gamma_\nu} \ I_{or} \exp(qV/2kT) \ . \tag{9}$$

It should be noted that the stress that alters the generation-recombination current is that which is located in the space charge region. This is different from that of the "ideal" currents where it is the stress at the edges of the space charge region that is important.

In the case of large reverse bias such that $qV/kT \ll 1$, I_r is given by

$$I_r = - \frac{qn_i^2 W}{\tau_{no}p_1 + \tau_{po}n_1} \ . \tag{10}$$

Assuming for the moment that $|E_t - E_i|$ is independent of stress, I_r reduces to

$$I_r = I_{or}' \sqrt{\gamma_\nu}(V + V_o)^{1/2} \ , \tag{11}$$

where I_{or}' is a constant. It is unlikely that $|E_t - E_i|$ is in fact constant. On the other hand, it is unlikely that the trap energy shifts would change in such a manner as to compensate very much for the changes in n_i.

III. COMPARISON WITH EXPERIMENT

In actual p-n junctions, the total current is the sum of the "ideal" and the generation-recombination. (Surface generation-recombination currents can be important, however, the preceding model is good only for bulk generation-recombination.) The total current is then

$$I_T = I_I + I_r \ . \tag{12}$$

Consider a p-n junction in which the total junction is stressed. The "ideal" current is, from Eq. (14) of I,

$$I_I = I_{oI} \ \gamma_\nu \left(e^{qV/kT} - 1\right) \ , \tag{13}$$

where I_{oI} is the saturation current for zero applied stress. Adding the generation-recombination current to the ideal current for the forward biased case where Eq. (7) applies, gives

$$I_T = I_{oI} \ \gamma_\nu \ \exp(qV/kT) + I_{or} \ \sqrt{\gamma_\nu} \ \exp(qV/2kT) \ . \tag{14}$$

In germanium, diodes I_I usually dominates I_T while in silicon, I_r usually dominates I_T for biases less than a few tenths of a volt. From Eq. (14), one would expect the "ideal" component of current to be increased by stress by the factor $\sqrt{\gamma_\nu}$ more than is the generation-recombination component.

Silicon mesa diodes have been fabricated and tested to determine the effects of stress on the forward biased I-V characteristics. Diodes were formed by diffusing a p-type layer into an n-type substrate. The p-type layer was then etched away except for circular dots which defined the mesa diodes. The forward bias characteristics were dominated by generation-recombination currents at voltages below a few tenths of a volt and switched to "ideal" currents at higher voltages. The mesa diodes were stressed by applying a known force through a steel post to the mesa.

Figure 1 shows a typical plot of the characteristics of a diode formed on a 0.01 Ω-cm, (110) oriented substrate. The diameter of the mesa was 33 microns. The magnitudes of the applied forces are shown in

the figure. The solid lines were calculated using Eq. (14) with $I_{or} \sqrt{\gamma_{\nu}}$ and $I_{oI} \gamma_{\nu}$ chosen to give the best fit to the data. As shown in the figure, Eq. (14) gives a reasonable fit to the data. As predicted by the model, stress causes a larger change in the "ideal" current than it does in the generation-recombination current. This is better illustrated in Figure 2 where the experimental values of $I_{oI} \gamma_{\nu}$ and $I_{or} \sqrt{\gamma_{\nu}}$ are plotted as a function of applied force. As expected, the slope of the generation-recombination current is one-half that of the "ideal" current when plotted as a function of force on semilog paper. Note also the straight line in this type of plot. This shows that, in fact, $\gamma_{\nu} = C_1 \exp C_2 F/A$. The constant C_2 as determined from Figure 2 is larger than that predicted by theory. This probably results from uneven stressing due to surface irregularities of the mesa.

Due to the physical construction of the present diodes (no junction passivation), the surface contributions to the total current were dominant in the reverse biased condition. Although stress did, in fact, induce changes in the I-V characteristics for reverse biases, the characteristics and the stress induced changes were non-reproducible from sample to sample.

Based on data obtained by Matukura[10] on planar type silicon p-n junctions operated in both the forward and reverse biased condition, it appears that there is little if any change induced in the factor $|E_t - E_i|$ by stress.

IV. DISCUSSION

The model presented here has assumed that there are no reversible stress induced generation-recombination centers. At high dynamic stress levels, one often finds that the effects are not completely reversible whereas they are reversible for static loading. These hysteresis effects are usually small and negligible. A more common effect is a permanent or semi-permanent change in the I-V characteristics induced by extremely large stresses (stresses near or above the fracture strength of the semiconductor). These effects are attributable to stress induced dislocations which result in new generation-recombination centers. For example, we have observed that p-n junctions which exhibited "ideal" current at a given forward bias exhibited generation-recombination currents which were orders of magnitude larger than the "ideal" current after large stresses were applied and removed. The large stress induced generation-recombination currents were affected by stress in the same manner as any other ordinary generation-recombination currents. The current at a constant bias and constant stress did decrease with time, however, which indicates that the individual centers were annealing with time.

In order to compare experimental data with the present theory for generation-recombination currents and the theory presented in I for "ideal" currents, care must be used to separate the two components. Also, the contributions due to surface generation-recombination must be removed since neither theory holds for the surface currents.

Footnotes and References

[1] W. Bernard, W. Rindner, and H. Roth, J. Appl. Phys. 35, 1860 (1964).

[2] W. Rindner, and I. Braun, J. Appl. Phys. 34, 1958 (1963).

[3] J. Wortman, J. Hauser, and R. Burger, J. Appl. Phys. 35, 2122 (1964).

[4] W. Rindner, J. Appl. Phys. 36, 2513 (1965).

[5] W. Shockley and W. Read, Phys. Rev. 87, 835 (1952).

[6] R. Hall, Phys. Rev. 87, 387, (1952).

[7] C. Sah, R. Noyce, and W. Shockley, Proc. IRE 45, 1228 (1957).

[8] J. Hauser, Proc. IEEE 53, 743 (1965).

[9] M. Holland and W. Paul, Phys. Rev. 128, 30 (1962).

M. Nathan and W. Paul, Phys. Rev. 128, 38 (1962).

M. Holland and W. Paul, Phys. Rev. 128, 43 (1962).

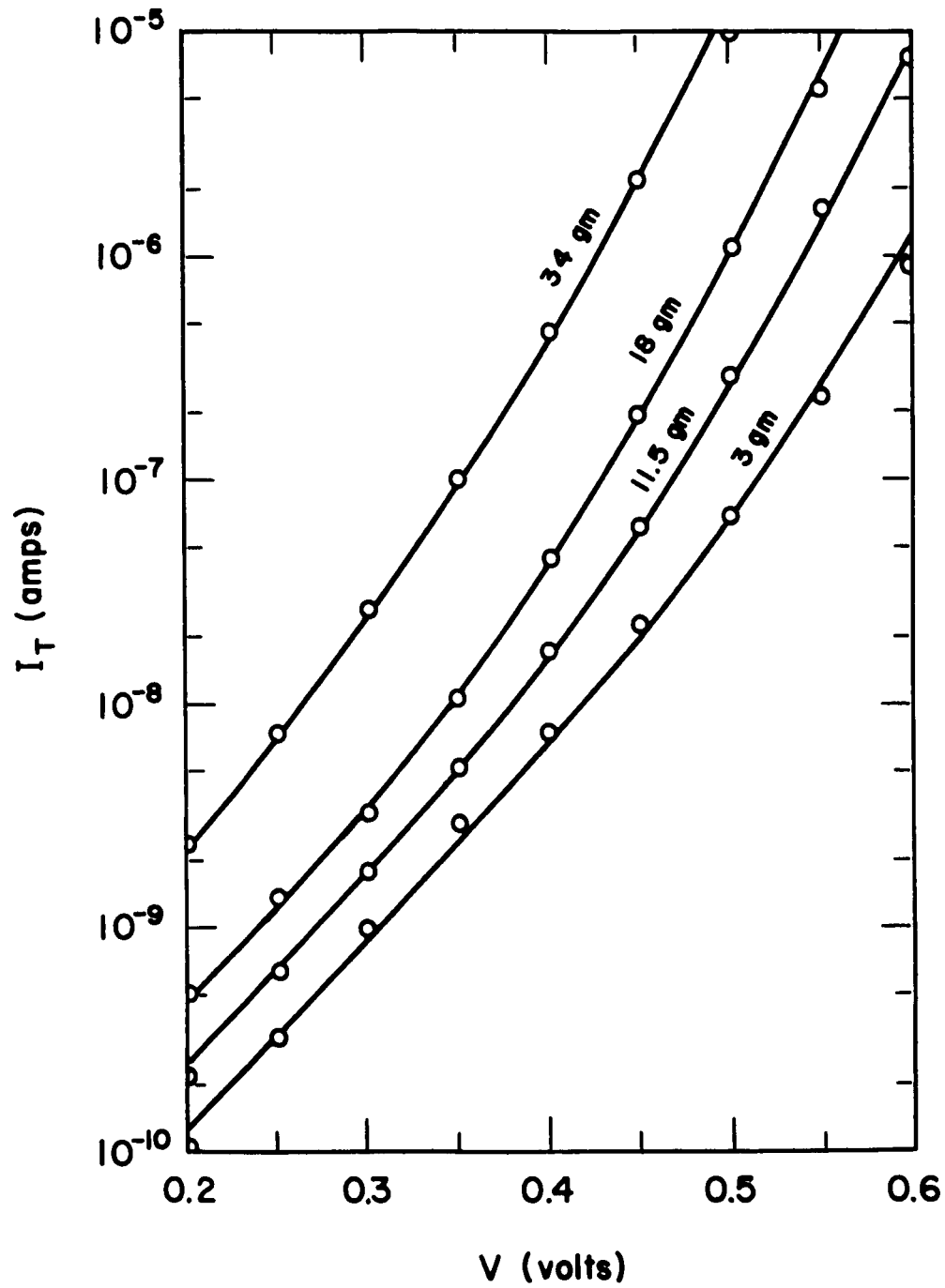[10] Y. Matukura, Japan J. Appl. Phys. 3, 256 (1964).

Figure 1. Forward Biased I-V characteristics of a mesa diode sub-
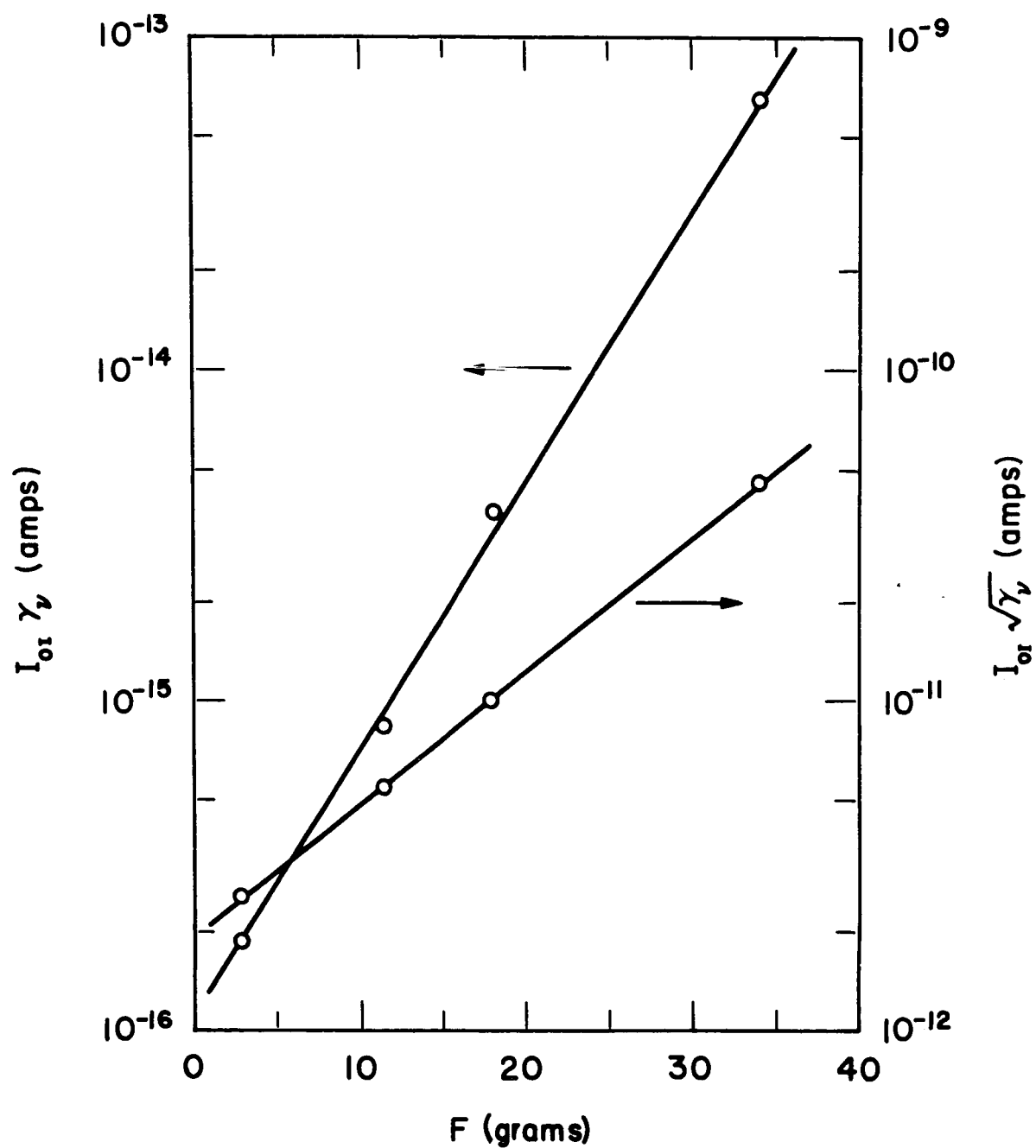jected to stress. (The applied force is shown.)

Figure 2. Experimental values of $I_{or} \sqrt{\gamma_\nu}$ and $I_{oI} \gamma_\nu$ as a function of applied force for the characteristics shown in Fig. 1.